Machine Learning with Python

MTH786U/P 2023/24

Detailed solutions Coursework 3

Nicola Perra, Queen Mary University of London (QMUL)

n.perra@qmul.ac.uk

Problem 1

Problem 1. Below you are asked to prove several small facts about convexity leading to a prove of the MSE function being convex.

1. Show that the sum of two convex functions is convex. **Hint**: use the definition of convexity.

2. Prove that, for any convex function $g : \mathcal{C} \subset \mathbb{R} \to \mathbb{R}$, the function $f(x) := ag(x) + b$ is also convex. Here $b \in \mathbb{R}$ is a scalar, and $a \in \mathbb{R}_+$ is a positive scalar (i.e. $a > 0$).

3. Verify that the function $h(w) := xw - y$ for fixed $x \in \mathbb{R}$ and $y \in \mathbb{R}$ satisfies

$$h(\lambda w + (1-\lambda)v) = \lambda h(w) + (1-\lambda)h(v)\,,$$

   for all $w, v \in \mathbb{R}$ and $\lambda \in [0, 1]$.

4. Show that the function $f(w) := g(h(w))$, where $g : \mathbb{R} \to \mathbb{R}$ is some convex function and h the function from Question 3, is convex.

5. Verify that the function $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with $g(x) := \frac{1}{2}x^2$ is convex.

6. Show that the simplified MSE function $\mathrm{MSE} : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with

$$\mathrm{MSE}(w) = \frac{1}{2}(xw - y)^2$$

   is convex.

   Hint: make us of Questions 1–5.

7. Prove that the general MSE function $\mathrm{MSE} : \mathbb{R}^{d+1} \to \mathbb{R}_{\geq 0}$ with

$$\mathrm{MSE}(\mathbf{w}) := \frac{1}{2s}\|\mathbf{X}\mathbf{w} - \mathbf{y}\|^2\,,$$

   for a matrix $\mathbf{X} \in \mathbb{R}^{s \times (d+1)}$ and a vector $\mathbf{y} \in \mathbb{R}^s$, is convex.

Reminder

A function $f: C \to \mathbb{R}$ over a convex set C is called *convex* if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

is satisfied for all $x, y \in C$ and $\lambda \in [0,1]$.

Reminder

A function $f : C \to \mathbb{R}$ over a convex set C is called *convex* if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

is satisfied for all $x, y \in C$ and $\lambda \in [0,1]$.

This definition assumes any property of the function f

Reminder

A twice differentiable function $f : C \to \mathbb{R}$ over a convex set C is called *convex* if

$$\frac{d^2 f}{dx^2} \geq 0$$

is satisfied for all $x \in C$.

Reminder

A twice differentiable function $f : C \to \mathbb{R}$ over a convex set C is called *convex* if

$$\frac{d^2 f}{dx^2} \geq 0$$

is satisfied for all $x \in C$.

For a function of n variables the condition is on the Hessian which should be positive semi-definite

1. Show that the sum of two convex functions is convex. **Hint**: use the definition of convexity.

1. Show that the sum of two convex functions is convex. **Hint**: use the definition of convexity.

We want to show that the sum of two convex functions is convex as well.

Let $f, g, h\colon \mathcal{C} \to \mathbb{R}$ such that for all $x \in \mathcal{C}$ we have $h(x) = f(x) + g(x)$, for two convex functions f and g. Then we observe the following:

1. Show that the sum of two convex functions is convex. **Hint**: use the definition of convexity.

We want to show that the sum of two convex functions is convex as well. Let $f, g, h \colon \mathcal{C} \to \mathbb{R}$ such that for all $x \in \mathcal{C}$ we have $h(x) = f(x) + g(x)$, for two convex functions f and g. Then we observe the following:

$$h(\lambda x + (1 - \lambda)y) =$$

1. Show that the sum of two convex functions is convex. **Hint**: use the definition of convexity.

We want to show that the sum of two convex functions is convex as well.

Let $f, g, h \colon \mathcal{C} \to \mathbb{R}$ such that for all $x \in \mathcal{C}$ we have $h(x) = f(x) + g(x)$, for two convex functions f and g. Then we observe the following:

$$h(\lambda x + (1-\lambda)y) = f(\lambda x + (1-\lambda)y) + g(\lambda x + (1-\lambda)y)$$

1. Show that the sum of two convex functions is convex. **Hint**: use the definition of convexity.

We want to show that the sum of two convex functions is convex as well.

Let $f, g, h\colon \mathcal{C} \to \mathbb{R}$ such that for all $x \in \mathcal{C}$ we have $h(x) = f(x) + g(x)$, for two convex functions f and g. Then we observe the following:

$$
\begin{aligned}
h(\lambda x + (1-\lambda)y) &= f(\lambda x + (1-\lambda)y) + g(\lambda x + (1-\lambda)y) \\
&\leq \lambda f(x) + (1-\lambda) f(y) + \lambda g(x) + (1-\lambda) g(y)
\end{aligned}
$$

1. Show that the sum of two convex functions is convex. **Hint**: use the definition of convexity.

We want to show that the sum of two convex functions is convex as well.

Let $f, g, h \colon \mathcal{C} \to \mathbb{R}$ such that for all $x \in \mathcal{C}$ we have $h(x) = f(x) + g(x)$, for two convex functions f and g. Then we observe the following:

$$
\begin{aligned}
h(\lambda x + (1-\lambda)y) &= f(\lambda x + (1-\lambda)y) + g(\lambda x + (1-\lambda)y) \\
&\leq \lambda f(x) + (1-\lambda) f(y) + \lambda g(x) + (1-\lambda) g(y) \\
&= \lambda \left[f(x) + g(x)\right] + (1-\lambda) \left[f(y) + g(y)\right]
\end{aligned}
$$

1. Show that the sum of two convex functions is convex. **Hint**: use the definition of convexity.

We want to show that the sum of two convex functions is convex as well.

Let $f, g, h\colon \mathcal{C} \to \mathbb{R}$ such that for all $x \in \mathcal{C}$ we have $h(x) = f(x) + g(x)$, for two convex functions f and g. Then we observe the following:

$$\begin{aligned} h(\lambda x + (1-\lambda)y) &= f(\lambda x + (1-\lambda)y) + g(\lambda x + (1-\lambda)y) \\ &\leq \lambda f(x) + (1-\lambda) f(y) + \lambda g(x) + (1-\lambda) g(y) \\ &= \lambda\left[f(x) + g(x)\right] + (1-\lambda)\left[f(y) + g(y)\right] \\ &= \lambda h(x) + (1-\lambda) h(y) \end{aligned}$$

2. Prove that, for any convex function $g : \mathcal{C} \subset \mathbb{R} \to \mathbb{R}$, the function $f(x) := ag(x) + b$ is also convex. Here $b \in \mathbb{R}$ is a scalar, and $a \in \mathbb{R}_+$ is a positive scalar (i.e. $a > 0$).

Again, we use the definition of convexity and show

$$f(\lambda x + (1 - \lambda)y) =$$

2. Prove that, for any convex function $g : \mathcal{C} \subset \mathbb{R} \to \mathbb{R}$, the function $f(x) := ag(x) + b$ is also convex. Here $b \in \mathbb{R}$ is a scalar, and $a \in \mathbb{R}_+$ is a positive scalar (i.e. $a > 0$).

Again, we use the definition of convexity and show

$$f(\lambda x + (1 - \lambda)y) = ag(\lambda x + (1 - \lambda)y) + b$$

2. Prove that, for any convex function $g : \mathcal{C} \subset \mathbb{R} \to \mathbb{R}$, the function $f(x) := ag(x) + b$ is also convex. Here $b \in \mathbb{R}$ is a scalar, and $a \in \mathbb{R}_+$ is a positive scalar (i.e. $a > 0$).

Again, we use the definition of convexity and show

$$
\begin{aligned}
f(\lambda x + (1-\lambda)y) &= ag(\lambda x + (1-\lambda)y) + b \\
&\leq a\lambda g(x) + a(1-\lambda)g(y) + b
\end{aligned}
$$

2. Prove that, for any convex function $g : \mathcal{C} \subset \mathbb{R} \to \mathbb{R}$, the function $f(x) := ag(x) + b$ is also convex. Here $b \in \mathbb{R}$ is a scalar, and $a \in \mathbb{R}_+$ is a positive scalar (i.e. $a > 0$).

Again, we use the definition of convexity and show

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= ag(\lambda x + (1-\lambda)y) + b \\ &\leq a\lambda g(x) + a(1-\lambda)g(y) + b \\ &= a\lambda g(x) + a(1-\lambda)g(y) + \lambda b + (1-\lambda)b \end{aligned}$$

2. Prove that, for any convex function $g : \mathcal{C} \subset \mathbb{R} \to \mathbb{R}$, the function $f(x) := ag(x) + b$ is also convex. Here $b \in \mathbb{R}$ is a scalar, and $a \in \mathbb{R}_+$ is a positive scalar (i.e. $a > 0$).

Again, we use the definition of convexity and show

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= ag(\lambda x + (1-\lambda)y) + b \\ &\leq a\lambda g(x) + a(1-\lambda)g(y) + b \\ &= a\lambda g(x) + a(1-\lambda)g(y) + \lambda b + (1-\lambda)b \\ &= \lambda\left(ag(x) + b\right) + (1-\lambda)\left(ag(y) + b\right) \end{aligned}$$

2. Prove that, for any convex function $g : \mathcal{C} \subset \mathbb{R} \to \mathbb{R}$, the function $f(x) := ag(x) + b$ is also convex. Here $b \in \mathbb{R}$ is a scalar, and $a \in \mathbb{R}_+$ is a positive scalar (i.e. $a > 0$).

Again, we use the definition of convexity and show

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= ag(\lambda x + (1-\lambda)y) + b \\ &\leq a\lambda g(x) + a(1-\lambda)g(y) + b \\ &= a\lambda g(x) + a(1-\lambda)g(y) + \lambda b + (1-\lambda)b \\ &= \lambda\left(ag(x) + b\right) + (1-\lambda)\left(ag(y) + b\right) \\ &= \lambda f(x) + (1-\lambda)f(y)\,, \end{aligned}$$

2. Prove that, for any convex function $g : \mathcal{C} \subset \mathbb{R} \to \mathbb{R}$, the function $f(x) := ag(x) + b$ is also convex. Here $b \in \mathbb{R}$ is a scalar, and $a \in \mathbb{R}_+$ is a positive scalar (i.e. $a > 0$).

Again, we use the definition of convexity and show

$$
\begin{aligned}
f(\lambda x + (1-\lambda)y) &= ag(\lambda x + (1-\lambda)y) + b \\
&\leq a\lambda g(x) + a(1-\lambda)g(y) + b \\
&= a\lambda g(x) + a(1-\lambda)g(y) + \lambda b + (1-\lambda)b \\
&= \lambda\left(ag(x) + b\right) + (1-\lambda)\left(ag(y) + b\right) \\
&= \lambda f(x) + (1-\lambda)f(y)\,,
\end{aligned}
$$

for all $x, y \in \mathcal{C}$ and $\lambda \in [0,1]$.

3. Verify that the function $h(w) := xw - y$ for fixed $x \in \mathbb{R}$ and $y \in \mathbb{R}$ satisfies

$$h(\lambda w + (1 - \lambda)v) = \lambda h(w) + (1 - \lambda)h(v)\,,$$

for all $w, v \in \mathbb{R}$ and $\lambda \in [0, 1]$.

3. Verify that the function $h(w) := xw - y$ for fixed $x \in \mathbb{R}$ and $y \in \mathbb{R}$ satisfies

$$h(\lambda w + (1-\lambda)v) = \lambda h(w) + (1-\lambda)h(v)\,,$$

for all $w, v \in \mathbb{R}$ and $\lambda \in [0,1]$.

$$h\left(\lambda w + (1-\lambda)v\right) =$$

3. Verify that the function $h(w) := xw - y$ for fixed $x \in \mathbb{R}$ and $y \in \mathbb{R}$ satisfies

$$h(\lambda w + (1-\lambda)v) = \lambda h(w) + (1-\lambda)h(v)\,,$$

for all $w, v \in \mathbb{R}$ and $\lambda \in [0,1]$.

$$h\left(\lambda w + (1-\lambda)v\right) = \lambda xw + (1-\lambda)vx - y$$

3. Verify that the function $h(w) := xw - y$ for fixed $x \in \mathbb{R}$ and $y \in \mathbb{R}$ satisfies

$$h(\lambda w + (1-\lambda)v) = \lambda h(w) + (1-\lambda)h(v)\,,$$

for all $w, v \in \mathbb{R}$ and $\lambda \in [0,1]$.

$$\begin{aligned} h\left(\lambda w + (1-\lambda)v\right) &= \lambda x w + (1-\lambda)vx - y \\ &= \lambda x w + (1-\lambda)vx - \lambda y - (1-\lambda)y \end{aligned}$$

3. Verify that the function $h(w) := xw - y$ for fixed $x \in \mathbb{R}$ and $y \in \mathbb{R}$ satisfies

$$h(\lambda w + (1-\lambda)v) = \lambda h(w) + (1-\lambda)h(v)\,,$$

for all $w, v \in \mathbb{R}$ and $\lambda \in [0,1]$.

$$\begin{aligned} h\left(\lambda w + (1-\lambda)v\right) &= \lambda xw + (1-\lambda)vx - y \\ &= \lambda xw + (1-\lambda)vx - \lambda y - (1-\lambda)y \\ &= \lambda(xw - y) + (1-\lambda)(vx - y) \end{aligned}$$

3. Verify that the function $h(w) := xw - y$ for fixed $x \in \mathbb{R}$ and $y \in \mathbb{R}$ satisfies

$$h(\lambda w + (1-\lambda)v) = \lambda h(w) + (1-\lambda)h(v)\,,$$

for all $w, v \in \mathbb{R}$ and $\lambda \in [0,1]$.

$$\begin{aligned} h\left(\lambda w + (1-\lambda)v\right) &= \lambda xw + (1-\lambda)vx - y \\ &= \lambda xw + (1-\lambda)vx - \lambda y - (1-\lambda)y \\ &= \lambda(xw - y) + (1-\lambda)(vx - y) \\ &= \lambda h(w) + (1-\lambda)h(v) \end{aligned}$$

4. Show that the function $f(w) := g(h(w))$, where $g : \mathbb{R} \to \mathbb{R}$ is some convex function and h the function from Question 3, is convex.

4. Show that the function $f(w) := g(h(w))$, where $g : \mathbb{R} \to \mathbb{R}$ is some convex function and h the function from Question 3, is convex.

For any convex function g and the function h from Exercise 3 we estimate

4. Show that the function $f(w) := g(h(w))$, where $g : \mathbb{R} \to \mathbb{R}$ is some convex function and h the function from Question 3, is convex.

For any convex function g and the function h from Exercise 3 we estimate

$$f(\lambda w + (1 - \lambda)v) =$$

4. Show that the function $f(w) := g(h(w))$, where $g : \mathbb{R} \to \mathbb{R}$ is some convex function and h the function from Question 3, is convex.

For any convex function g and the function h from Exercise 3 we estimate

$$f(\lambda w + (1 - \lambda)v) = g(h(\lambda w + (1 - \lambda)v))$$

4. Show that the function $f(w) := g(h(w))$, where $g : \mathbb{R} \to \mathbb{R}$ is some convex function and h the function from Question 3, is convex.

For any convex function g and the function h from Exercise 3 we estimate

$$\begin{aligned} f(\lambda w + (1-\lambda)v) &= g(h(\lambda w + (1-\lambda)v)) \\ &= g(\lambda h(w) + (1-\lambda)h(v)) \end{aligned}$$

4. Show that the function $f(w) := g(h(w))$, where $g : \mathbb{R} \to \mathbb{R}$ is some convex function and h the function from Question 3, is convex.

For any convex function g and the function h from Exercise 3 we estimate

$$\begin{aligned} f(\lambda w + (1-\lambda)v) &= g(h(\lambda w + (1-\lambda)v)) \\ &= g(\lambda h(w) + (1-\lambda)h(v)) \\ &\leq \lambda g(h(w)) + (1-\lambda)g(h(v)) \end{aligned}$$

4. Show that the function $f(w) := g(h(w))$, where $g : \mathbb{R} \to \mathbb{R}$ is some convex function and h the function from Question 3, is convex.

For any convex function g and the function h from Exercise 3 we estimate

$$\begin{aligned} f(\lambda w + (1-\lambda)v) &= g(h(\lambda w + (1-\lambda)v)) \\ &= g(\lambda h(w) + (1-\lambda)h(v)) \\ &\leq \lambda g(h(w)) + (1-\lambda)g(h(v)) \\ &= \lambda f(w) + (1-\lambda)f(v)\,. \end{aligned}$$

4. Show that the function $f(w) := g(h(w))$, where $g : \mathbb{R} \to \mathbb{R}$ is some convex function and h the function from Question 3, is convex.

For any convex function g and the function h from Exercise 3 we estimate

$$\begin{aligned} f(\lambda w + (1-\lambda)v) &= g(h(\lambda w + (1-\lambda)v)) \\ &= g(\lambda h(w) + (1-\lambda)h(v)) \\ &\leq \lambda g(h(w)) + (1-\lambda)g(h(v)) \\ &= \lambda f(w) + (1-\lambda)f(v)\,. \end{aligned}$$

Thus, the composition $g(h(w))$ is also convex.

5. Verify that the function $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with $g(x) := \frac{1}{2}x^2$ is convex.

5. Verify that the function $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with $g(x) := \frac{1}{2}x^2$ is convex.

For the function $g(x) := \frac{1}{2}x^2$ we estimate

5. Verify that the function $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with $g(x) := \frac{1}{2}x^2$ is convex.

For the function $g(x) := \frac{1}{2}x^2$ we estimate

$$2\lambda g(x) + 2(1-\lambda)g(y) - 2g(\lambda x + (1-\lambda)y)$$

5. Verify that the function $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with $g(x) := \frac{1}{2}x^2$ is convex.

For the function $g(x) := \frac{1}{2}x^2$ we estimate

$$\begin{aligned} &2\lambda g(x) + 2(1-\lambda)g(y) - 2g(\lambda x + (1-\lambda)y) \\ = &\lambda x^2 + (1-\lambda)y^2 - (\lambda x + (1-\lambda)y)^2 \end{aligned}$$

5. Verify that the function $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with $g(x) := \frac{1}{2}x^2$ is convex.

For the function $g(x) := \frac{1}{2}x^2$ we estimate

$$
\begin{aligned}
&2\lambda g(x) + 2(1-\lambda)g(y) - 2g(\lambda x + (1-\lambda)y) \\
&= \lambda x^2 + (1-\lambda)y^2 - (\lambda x + (1-\lambda)y)^2 \\
&= \lambda x^2 + (1-\lambda)y^2 - \lambda^2 x^2 - 2\lambda(1-\lambda)xy - (1-\lambda)^2 y^2
\end{aligned}
$$

5. Verify that the function $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with $g(x) := \frac{1}{2}x^2$ is convex.

For the function $g(x) := \frac{1}{2}x^2$ we estimate

$$
\begin{aligned}
& 2\lambda g(x) + 2(1-\lambda)g(y) - 2g(\lambda x + (1-\lambda)y) \\
&= \lambda x^2 + (1-\lambda)y^2 - (\lambda x + (1-\lambda)y)^2 \\
&= \lambda x^2 + (1-\lambda)y^2 - \lambda^2 x^2 - 2\lambda(1-\lambda)xy - (1-\lambda)^2 y^2 \\
&= \lambda(1-\lambda)x^2 + \lambda(1-\lambda)y^2 - 2\lambda(1-\lambda)xy
\end{aligned}
$$

5. Verify that the function $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with $g(x) := \frac{1}{2}x^2$ is convex.

For the function $g(x) := \frac{1}{2}x^2$ we estimate

$$\begin{aligned}
&2\lambda g(x) + 2(1-\lambda)g(y) - 2g(\lambda x + (1-\lambda)y) \\
&= \lambda x^2 + (1-\lambda)y^2 - (\lambda x + (1-\lambda)y)^2 \\
&= \lambda x^2 + (1-\lambda)y^2 - \lambda^2 x^2 - 2\lambda(1-\lambda)xy - (1-\lambda)^2 y^2 \\
&= \lambda(1-\lambda)x^2 + \lambda(1-\lambda)y^2 - 2\lambda(1-\lambda)xy \\
&= \lambda(1-\lambda)(x-y)^2 \geq 0\,,
\end{aligned}$$

5. Verify that the function $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with $g(x) := \frac{1}{2}x^2$ is convex.

For the function $g(x) := \frac{1}{2}x^2$ we estimate

$$\begin{aligned}
&2\lambda g(x) + 2(1-\lambda)g(y) - 2g(\lambda x + (1-\lambda)y) \\
&= \lambda x^2 + (1-\lambda)y^2 - (\lambda x + (1-\lambda)y)^2 \\
&= \lambda x^2 + (1-\lambda)y^2 - \lambda^2 x^2 - 2\lambda(1-\lambda)xy - (1-\lambda)^2 y^2 \\
&= \lambda(1-\lambda)x^2 + \lambda(1-\lambda)y^2 - 2\lambda(1-\lambda)xy \\
&= \lambda(1-\lambda)(x-y)^2 \geq 0\,,
\end{aligned}$$

since $\lambda(1-\lambda) \geq 0$ for $\lambda \in [0,1]$, which implies

$$g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y)\,.$$

Hence, we have concluded that g is convex.

5. Verify that the function $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with $g(x) := \frac{1}{2}x^2$ is convex.

5. Verify that the function $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with $g(x) := \frac{1}{2}x^2$ is convex.

Alternatively: we know that the function is twice differentiable

5. Verify that the function $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with $g(x) := \frac{1}{2}x^2$ is convex.

Alternatively: we know that the function is twice differentiable

$$\frac{d}{dx}\left(\frac{1}{2}x^2\right) = x$$

5. Verify that the function $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with $g(x) := \frac{1}{2}x^2$ is convex.

Alternatively: we know that the function is twice differentiable

$$\frac{d}{dx}\left(\frac{1}{2}x^2\right) = x$$

$$\frac{d^2}{dx^2}\left(\frac{1}{2}x^2\right) = 1 \geq 0$$

6. Show that the simplified MSE function $\text{MSE} : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with

$$\text{MSE}(w) = \frac{1}{2}(xw - y)^2$$

is convex.

Hint: make us of Questions 1–5.

6. Show that the simplified MSE function $\text{MSE} : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with

$$\text{MSE}(w) = \frac{1}{2}(xw - y)^2$$

is convex.

Hint: make us of Questions 1–5.

We verify this result by combining the results from Exercise 3, Exercise 4 and Exercise 5. We can write $\text{MSE}(w) = g(h(w))$, for $h(w) := xw - y$ and $g(z) := \frac{1}{2}z^2$. From Exercise 5 we know that g is convex and from Exercise 4 we know that the composition $g \circ h$ is convex. Since this is equivalent to the MSE, we already know that the MSE is convex.

7. Prove that the general MSE function $\mathrm{MSE} : \mathbb{R}^{d+1} \to \mathbb{R}_{\geq 0}$ with

$$\mathrm{MSE}(\mathbf{w}) := \frac{1}{2s}\|\mathbf{X}\mathbf{w} - \mathbf{y}\|^2,$$

for a matrix $\mathbf{X} \in \mathbb{R}^{s\times(d+1)}$ and a vector $\mathbf{y} \in \mathbb{R}^s$, is convex.

7. Prove that the general MSE function $\mathrm{MSE} : \mathbb{R}^{d+1} \to \mathbb{R}_{\geq 0}$ with

$$\mathrm{MSE}(\mathbf{w}) := \frac{1}{2s}\|\mathbf{X}\mathbf{w} - \mathbf{y}\|^2,$$

for a matrix $\mathbf{X} \in \mathbb{R}^{s\times(d+1)}$ and a vector $\mathbf{y} \in \mathbb{R}^s$, is convex.

We proceed in similar fashion as in the previous exercise. We point out that the MSE can be written as MSE$(w) = g(h(w))$ for $g(y) = \frac{1}{2s}\|z\|^2 = \frac{1}{2s}\sum_{i=1}^{s}|z_i|^2$ and $h(w) = Xw - y$. Note that g is convex since the function $x \to x^2$ is convex (see Exercise 5) and since the sum of convex functions is also convex (see Exercise 1). In the same way as in Exercise 3 we verify

$$h(\lambda w + (1-\lambda)v) = \lambda h(w) + (1-\lambda)h(v);$$

hence, MSE is a composition of a convex and an affine-linear function and as a consequence of Exercise 4, MSE is convex.

Problem 2

Problem 2. Set up a linear regression problem of the form

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathbb{R}^2} \left\{ \frac{1}{2s} \sum_{i=1}^{3} |w^{(0)} + w^{(1)} x^{(i)} - y^{(i)}|^2 \right\} , \qquad (1)$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = -c$ and $y^{(1)} = 2$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 0$ and $y^{(2)} = 2$, and $(x^{(3)}, y^{(3)})$ with $x^{(3)} = c$ and $y^{(3)} = 2$, for some constant $c > 0$.

1. Derive the normal equation for this problem.
2. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.
3. Repeat the previous exercise, but this time assume you make an error in your measurement. The new, perturbed measurements $\mathbf{y}_\delta$ read $y_\delta^{(1)} = 2 + \varepsilon$, $y_\delta^{(2)} = 2 + \varepsilon$ and $y_\delta^{(3)} = 2 - \varepsilon$.
4. Compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm.
5. How does the error compare with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = -c$ and $y^{(1)} = 2$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 0$ and $y^{(2)} = 2$, and $(x^{(3)}, y^{(3)})$ with $x^{(3)} = c$ and $y^{(3)} = 2$, for some constant $c > 0$.

1. Derive the normal equation for this problem.

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = -c$ and $y^{(1)} = 2$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 0$ and $y^{(2)} = 2$, and $(x^{(3)}, y^{(3)})$ with $x^{(3)} = c$ and $y^{(3)} = 2$, for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X^\top X \hat{w} = X^\top y$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = -c$ and $y^{(1)} = 2$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 0$ and $y^{(2)} = 2$, and $(x^{(3)}, y^{(3)})$ with $x^{(3)} = c$ and $y^{(3)} = 2$, for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X^\top X\hat{w} = X^\top y$$

$$X = \begin{pmatrix} 1 & -c \\ 1 & 0 \\ 1 & c \end{pmatrix}$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = -c$ and $y^{(1)} = 2$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 0$ and $y^{(2)} = 2$, and $(x^{(3)}, y^{(3)})$ with $x^{(3)} = c$ and $y^{(3)} = 2$, for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X^\top X\hat{w} = X^\top y$$

$$X = \begin{pmatrix} 1 & -c \\ 1 & 0 \\ 1 & c \end{pmatrix}$$

$$X^\top = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix}$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = -c$ and $y^{(1)} = 2$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 0$ and $y^{(2)} = 2$, and $(x^{(3)}, y^{(3)})$ with $x^{(3)} = c$ and $y^{(3)} = 2$, for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X^\top X\hat{w} = X^\top y$$

$$X = \begin{pmatrix} 1 & -c \\ 1 & 0 \\ 1 & c \end{pmatrix} \qquad X^\top = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \qquad y = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = -c$ and $y^{(1)} = 2$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 0$ and $y^{(2)} = 2$, and $(x^{(3)}, y^{(3)})$ with $x^{(3)} = c$ and $y^{(3)} = 2$, for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X^\top X\hat{w} = X^\top y$$

$$X = \begin{pmatrix} 1 & -c \\ 1 & 0 \\ 1 & c \end{pmatrix} \qquad X^\top = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \qquad y = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$$

$$X^\top X\hat{w} = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \begin{pmatrix} 1 & -c \\ 1 & 0 \\ 1 & c \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix}$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = -c$ and $y^{(1)} = 2$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 0$ and $y^{(2)} = 2$, and $(x^{(3)}, y^{(3)})$ with $x^{(3)} = c$ and $y^{(3)} = 2$, for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X^\top X\hat{w} = X^\top y$$

$$X = \begin{pmatrix} 1 & -c \\ 1 & 0 \\ 1 & c \end{pmatrix} \qquad X^\top = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \qquad y = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$$

$$X^\top X\hat{w} = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \begin{pmatrix} 1 & -c \\ 1 & 0 \\ 1 & c \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix}$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = -c$ and $y^{(1)} = 2$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 0$ and $y^{(2)} = 2$, and $(x^{(3)}, y^{(3)})$ with $x^{(3)} = c$ and $y^{(3)} = 2$, for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X^\top = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \qquad y = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = -c$ and $y^{(1)} = 2$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 0$ and $y^{(2)} = 2$, and $(x^{(3)}, y^{(3)})$ with $x^{(3)} = c$ and $y^{(3)} = 2$, for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X^\top = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \qquad y = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = X^\top y = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = -c$ and $y^{(1)} = 2$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 0$ and $y^{(2)} = 2$, and $(x^{(3)}, y^{(3)})$ with $x^{(3)} = c$ and $y^{(3)} = 2$, for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X^\top = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \qquad y = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = X^\top y = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}$$

2. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}$$

2. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}$$

$$3w_0 = 6 \rightarrow w_0 = 2$$

2. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}$$

$$3w_0 = 6 \rightarrow w_0 = 2$$

$$2c^2 w_1 = 0 \rightarrow w_1 = 0$$

2. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}$$

$$3w_0 = 6 \rightarrow w_0 = 2$$

$$2c^2 w_1 = 0 \rightarrow w_1 = 0$$

$$\begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

3. Repeat the previous exercise, but this time assume you make an error in your measurement. The new, perturbed measurements $\mathbf{y}_\delta$ read $y_\delta^{(1)} = 2 + \varepsilon$, $y_\delta^{(2)} = 2 + \varepsilon$ and $y_\delta^{(3)} = 2 - \varepsilon$.

3. Repeat the previous exercise, but this time assume you make an error in your measurement. The new, perturbed measurements $\mathbf{y}_\delta$ read $y_\delta^{(1)} = 2 + \varepsilon$, $y_\delta^{(2)} = 2 + \varepsilon$ and $y_\delta^{(3)} = 2 - \varepsilon$.

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = X^\top y = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \begin{pmatrix} 2+\epsilon \\ 2+\epsilon \\ 2-\epsilon \end{pmatrix}$$

3. Repeat the previous exercise, but this time assume you make an error in your measurement. The new, perturbed measurements $\mathbf{y}_\delta$ read $y_\delta^{(1)} = 2 + \varepsilon$, $y_\delta^{(2)} = 2 + \varepsilon$ and $y_\delta^{(3)} = 2 - \varepsilon$.

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = X^\top y = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \begin{pmatrix} 2+\epsilon \\ 2+\epsilon \\ 2-\epsilon \end{pmatrix}$$

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 6+\epsilon \\ -2c\epsilon \end{pmatrix}$$

3. Repeat the previous exercise, but this time assume you make an error in your measurement. The new, perturbed measurements $\mathbf{y}_\delta$ read $y_\delta^{(1)} = 2 + \varepsilon$, $y_\delta^{(2)} = 2 + \varepsilon$ and $y_\delta^{(3)} = 2 - \varepsilon$.

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = X^\top y = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} \begin{pmatrix} 2 + \epsilon \\ 2 + \epsilon \\ 2 - \epsilon \end{pmatrix}$$

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 6 + \epsilon \\ -2c\epsilon \end{pmatrix}$$

$$\begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 + \frac{\epsilon}{3} \\ -\frac{\epsilon}{c} \end{pmatrix}$$

4. Compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm.

4. Compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm.

$$\|\hat{w} - \hat{w}_\delta\| = \sqrt{\sum_i (\hat{w}_i - \hat{w}_{\delta_i})^2}$$

4. Compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm.

$$\|\hat{w} - \hat{w}_\delta\| = \sqrt{\sum_i (\hat{w}_i - \hat{w}_{\delta_i})^2}$$

$$\hat{w} = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

4. Compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm.

$$\|\hat{w} - \hat{w}_\delta\| = \sqrt{\sum_i (\hat{w}_i - \hat{w}_{\delta_i})^2}$$

$$\hat{w} = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \qquad \hat{w}_\delta = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 + \frac{\epsilon}{3} \\ -\frac{\epsilon}{c} \end{pmatrix}$$

4. Compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm.

$$\|\hat{w} - \hat{w}_\delta\| = \sqrt{\sum_i (\hat{w}_i - \hat{w}_{\delta_i})^2}$$

$$\hat{w} = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \qquad \hat{w}_\delta = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 + \frac{\epsilon}{3} \\ -\frac{\epsilon}{c} \end{pmatrix}$$

$$\|\hat{w} - \hat{w}_\delta\| = \sqrt{\left(2 - \left(2 + \frac{\varepsilon}{3}\right)\right)^2 + \left(0 - \frac{\varepsilon}{c}\right)^2}$$

4. Compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm.

$$\|\hat{w} - \hat{w}_\delta\| = \sqrt{\sum_i (\hat{w}_i - \hat{w}_{\delta_i})^2}$$

$$\hat{w} = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \qquad \hat{w}_\delta = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 + \frac{\epsilon}{3} \\ -\frac{\epsilon}{c} \end{pmatrix}$$

$$\|\hat{w} - \hat{w}_\delta\| = \sqrt{\left(2 - \left(2 + \frac{\varepsilon}{3}\right)\right)^2 + \left(0 - \frac{\varepsilon}{c}\right)^2} = \sqrt{\frac{\varepsilon^2}{9} + \frac{\varepsilon^2}{c^2}} = \frac{\varepsilon\sqrt{9 + c^2}}{3c}$$

4. Compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm.

$$\|\hat{w} - \hat{w}_\delta\| = \sqrt{\sum_i (\hat{w}_i - \hat{w}_{\delta_i})^2}$$

$$\hat{w} = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \qquad \hat{w}_\delta = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 + \frac{\epsilon}{3} \\ -\frac{\epsilon}{c} \end{pmatrix}$$

$$\|\hat{w} - \hat{w}_\delta\| = \sqrt{\left(2 - \left(2 + \frac{\varepsilon}{3}\right)\right)^2 + \left(0 - \frac{\varepsilon}{c}\right)^2} = \sqrt{\frac{\varepsilon^2}{9} + \frac{\varepsilon^2}{c^2}} = \frac{\varepsilon\sqrt{9 + c^2}}{3c}$$
$$= \frac{\varepsilon}{c}\sqrt{1 + \left(\frac{c}{3}\right)^2} > \frac{\varepsilon}{c}.$$

5. How does the error compare with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?

5. How does the error compare with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?

$$\|\hat{w} - \hat{w}_\delta\| = \frac{\epsilon}{c}\sqrt{1 + \frac{c^2}{9}}$$

5. How does the error compare with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?

$$\|\hat{w} - \hat{w}_\delta\| = \frac{\epsilon}{c}\sqrt{1 + \frac{c^2}{9}}$$

$$\|y - y_\delta\| = \epsilon\sqrt{3}$$

5. How does the error compare with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?

$$\|\hat{w} - \hat{w}_\delta\| = \frac{\epsilon}{c}\sqrt{1 + \frac{c^2}{9}}$$

$$\|y - y_\delta\| = \epsilon\sqrt{3}$$

$$\|\hat{w} - \hat{w}_\delta\| \gg \| y - y_\delta\| \quad for \quad c \to 0$$

5. How does the error compare with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?

$$\|\hat{w} - \hat{w}_\delta\| = \frac{\epsilon}{c}\sqrt{1 + \frac{c^2}{9}}$$

$$\|y - y_\delta\| = \epsilon\sqrt{3}$$

$$\|\hat{w} - \hat{w}_\delta\| \gg \| y - y_\delta\| \;\; for \;\; c \to 0$$

The error in reconstruction is dominated by the ratio ε/c. If $c \ll \varepsilon$ the error can get potentially very large compared to the data error $\delta = \|y - y^\delta\| = \varepsilon\sqrt{3}$, which does not depend on c. Suppose $\varepsilon = 1/100$ and $c = 1/1000$, then $\delta \approx 0.01732$ but $\varepsilon/c = 10$. Hence, the data error δ is amplified by a factor larger than 577 in the reconstruction.

Problem 3

Problem 3. Let us consider a standard normal equation for a linear regression in dimensions $d \times 1$ (i.e. output is $n = 1$ dimensional). Let $\mathbf{y}$ and $\mathbf{y}_\delta$ be non-perturbed and perturbed output data correspondingly.

$$\|\hat{\mathbf{w}} - \hat{\mathbf{w}}_\delta\|^2 = \sum_{j=1}^{d+1} \sigma_j^{-2} \left|\langle \mathbf{u}^{(j)}, \mathbf{y} - \mathbf{y}_\delta \rangle\right|^2$$

for two least-squares solutions $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ with singular value decompositions

$$\hat{\mathbf{w}} = \sum_{j=1}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, \mathbf{y} \rangle \quad \text{and} \quad \hat{\mathbf{w}}_\delta = \sum_{j=1}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, \mathbf{y}_\delta \rangle ,$$

where $\sigma_j, \mathbf{u}^{(j)}, \mathbf{v}^{(j)}$ are singular values and right-/left- singular vectors of matrix $\mathbf{X}$. **Hint:** make use of the fact that singular vectors are orthonormal.

$$\|\hat{\mathbf{w}} - \hat{\mathbf{w}}_\delta\|^2 = \left\| \sum_{j=1}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \right\|^2$$

$$
\begin{aligned}
\|\hat{\mathbf{w}} - \hat{\mathbf{w}}_\delta\|^2 &= \left\| \sum_{j=1}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \right\|^2 \\
&= \left\| \sigma_1^{-1} \mathbf{v}^{(1)} \langle \mathbf{u}^{(1)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle + \sum_{j=2}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \right\|^2
\end{aligned}
$$

$$\|\hat{\mathbf{w}} - \hat{\mathbf{w}}_\delta\|^2 = \left\| \sum_{j=1}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \right\|^2$$

$$= \left\| \sigma_1^{-1} \mathbf{v}^{(1)} \langle \mathbf{u}^{(1)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle + \sum_{j=2}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \right\|^2$$

$$\|a + b\|^2 = \sum_i (a_i + b_i)^2$$

$$\|\hat{\mathbf{w}} - \hat{\mathbf{w}}_\delta\|^2 = \left\| \sum_{j=1}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \right\|^2$$

$$= \left\| \sigma_1^{-1} \mathbf{v}^{(1)} \langle \mathbf{u}^{(1)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle + \sum_{j=2}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \right\|^2$$

$$\|a + b\|^2 = \sum_i (a_i + b_i)^2$$

$$= \sum_i (a_i^2 + 2a_i b_i + b_i^2)$$

$$\|\hat{\mathbf{w}} - \hat{\mathbf{w}}_\delta\|^2 = \left\| \sum_{j=1}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \right\|^2$$

$$= \left\| \sigma_1^{-1} \mathbf{v}^{(1)} \langle \mathbf{u}^{(1)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle + \sum_{j=2}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \right\|^2$$

$$\|a + b\|^2 = \sum_i (a_i + b_i)^2$$

$$= \sum_i (a_i^2 + 2a_i b_i + b_i^2)$$

$$= \|a\|^2 + 2\langle a, b \rangle + \|b\|^2$$

$$= \left\| \sigma_1^{-1} \mathbf{v}^{(1)} \langle \mathbf{u}^{(1)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle + \sum_{j=2}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \right\|^2$$

$$= \left\| \sigma_1^{-1}\mathbf{v}^{(1)}\langle \mathbf{u}^{(1)}, (\mathbf{y}-\mathbf{y}_\delta)\rangle + \sum_{j=2}^{d+1} \sigma_j^{-1}\mathbf{v}^{(j)}\langle \mathbf{u}^{(j)}, (\mathbf{y}-\mathbf{y}_\delta)\rangle \right\|^2$$

$$\begin{aligned} &= \left\| \sigma_1^{-1}\mathbf{v}^{(1)}\langle \mathbf{u}^{(1)}, (\mathbf{y}-\mathbf{y}_\delta)\rangle \right\|^2 \\ &\quad - 2\sigma_1^{-1}\langle \mathbf{u}^{(1)}, (\mathbf{y}-\mathbf{y}_\delta)\rangle \left\langle \mathbf{v}^{(1)}, \sum_{j=2}^{d+1} \sigma_j^{-1}\mathbf{v}^{(j)}\langle \mathbf{u}^{(j)}, (\mathbf{y}-\mathbf{y}_\delta)\rangle \right\rangle \\ &\quad + \left\| \sum_{j=2}^{d+1} \sigma_j^{-1}\mathbf{v}^{(j)}\langle \mathbf{u}^{(j)}, (\mathbf{y}-\mathbf{y}_\delta)\rangle \right\|^2 \end{aligned}$$

$$= \left\|\sigma_1^{-1}\mathbf{v}^{(1)}\langle\mathbf{u}^{(1)},(\mathbf{y}-\mathbf{y}_\delta)\rangle\right\|^2$$

$$-2\sigma_1^{-1}\langle\mathbf{u}^{(1)},(\mathbf{y}-\mathbf{y}_\delta)\rangle\left\langle\mathbf{v}^{(1)},\sum_{j=2}^{d+1}\sigma_j^{-1}\mathbf{v}^{(j)}\langle\mathbf{u}^{(j)},(\mathbf{y}-\mathbf{y}_\delta)\rangle\right\rangle$$

$$+\left\|\sum_{j=2}^{d+1}\sigma_j^{-1}\mathbf{v}^{(j)}\langle\mathbf{u}^{(j)},(\mathbf{y}-\mathbf{y}_\delta)\rangle\right\|^2$$

$$= \sigma_1^{-2}\left|\langle\mathbf{u}^{(1)},(\mathbf{y}-\mathbf{y}_\delta)\rangle\right|^2\|\mathbf{v}^{(1)}\|^2$$

$$= \left\|\sigma_1^{-1}\mathbf{v}^{(1)}\langle\mathbf{u}^{(1)},(\mathbf{y}-\mathbf{y}_\delta)\rangle\right\|^2$$

$$-\, 2\sigma_1^{-1}\langle\mathbf{u}^{(1)},(\mathbf{y}-\mathbf{y}_\delta)\rangle\left\langle\mathbf{v}^{(1)},\sum_{j=2}^{d+1}\sigma_j^{-1}\mathbf{v}^{(j)}\langle\mathbf{u}^{(j)},(\mathbf{y}-\mathbf{y}_\delta)\rangle\right\rangle$$

$$+\left\|\sum_{j=2}^{d+1}\sigma_j^{-1}\mathbf{v}^{(j)}\langle\mathbf{u}^{(j)},(\mathbf{y}-\mathbf{y}_\delta)\rangle\right\|^2$$

$$= \sigma_1^{-2}\left|\langle\mathbf{u}^{(1)},(\mathbf{y}-\mathbf{y}_\delta)\rangle\right|^2\|\mathbf{v}^{(1)}\|^2$$

$$-\, 2\sigma_1^{-1}\langle\mathbf{u}^{(1)},(\mathbf{y}-\mathbf{y}_\delta)\rangle\sum_{j=2}^{d+1}\sigma_j^{-1}\left\langle\mathbf{v}^{(1)},\mathbf{v}^{(j)}\right\rangle\langle\mathbf{u}^{(j)},(\mathbf{y}-\mathbf{y}_\delta)\rangle$$

$$= \left\| \sigma_1^{-1} \mathbf{v}^{(1)} \langle \mathbf{u}^{(1)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \right\|^2$$

$$- 2\sigma_1^{-1} \langle \mathbf{u}^{(1)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \left\langle \mathbf{v}^{(1)}, \sum_{j=2}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \right\rangle$$

$$+ \left\| \sum_{j=2}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \right\|^2$$

$$= \sigma_1^{-2} \left| \langle \mathbf{u}^{(1)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \right|^2 \|\mathbf{v}^{(1)}\|^2$$

$$- 2\sigma_1^{-1} \langle \mathbf{u}^{(1)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \sum_{j=2}^{d+1} \sigma_j^{-1} \left\langle \mathbf{v}^{(1)}, \mathbf{v}^{(j)} \right\rangle \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle$$

$$+ \left\| \sum_{j=2}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \right\|^2$$

$$
\begin{aligned}
&= \sigma_1^{-2} \left| \langle \mathbf{u}^{(1)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \right|^2 \| \mathbf{v}^{(1)} \|^2 \\
&\quad - 2\sigma_1^{-1} \langle \mathbf{u}^{(1)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \sum_{j=2}^{d+1} \sigma_j^{-1} \left\langle \mathbf{v}^{(1)}, \mathbf{v}^{(j)} \right\rangle \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \\
&\quad + \left\| \sum_{j=2}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \right\|^2
\end{aligned}
$$

$$= \sigma_1^{-2} \left|\langle \mathbf{u}^{(1)}, (\mathbf{y} - \mathbf{y}_\delta)\rangle\right|^2 \|\mathbf{v}^{(1)}\|^2$$

1

$$- 2\sigma_1^{-1}\langle \mathbf{u}^{(1)}, (\mathbf{y} - \mathbf{y}_\delta)\rangle \sum_{j=2}^{d+1} \sigma_j^{-1} \left\langle \mathbf{v}^{(1)}, \mathbf{v}^{(j)} \right\rangle \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta)\rangle$$

$$+ \left\| \sum_{j=2}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta)\rangle \right\|^2$$

$$
\begin{aligned}
&= \sigma_1^{-2} \left|\langle \mathbf{u}^{(1)}, (\mathbf{y} - \mathbf{y}_\delta)\rangle\right|^2 \|\mathbf{v}^{(1)}\|^2 \\
&\quad - 2\sigma_1^{-1} \langle \mathbf{u}^{(1)}, (\mathbf{y} - \mathbf{y}_\delta)\rangle \sum_{j=2}^{d+1} \sigma_j^{-1} \langle \mathbf{v}^{(1)}, \mathbf{v}^{(j)} \rangle \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta)\rangle \\
&\quad + \left\| \sum_{j=2}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta)\rangle \right\|^2
\end{aligned}
$$

1

0

$$= \sigma_1^{-2} \left|\langle \mathbf{u}^{(1)}, (\mathbf{y} - \mathbf{y}_\delta)\rangle\right|^2 \|\mathbf{v}^{(1)}\|^2$$

$$- 2\sigma_1^{-1}\langle \mathbf{u}^{(1)}, (\mathbf{y} - \mathbf{y}_\delta)\rangle \sum_{j=2}^{d+1} \sigma_j^{-1} \langle \mathbf{v}^{(1)}, \mathbf{v}^{(j)}\rangle \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta)\rangle$$

$$+ \left\| \sum_{j=2}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta)\rangle \right\|^2$$

1

0

$$= \sigma_1^{-2} \left|\langle \mathbf{u}^{(1)}, (\mathbf{y} - \mathbf{y}_\delta)\rangle\right|^2 + \left\| \sum_{j=2}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta)\rangle \right\|^2 .$$

$$\|\hat{\mathbf{w}} - \hat{\mathbf{w}}_\delta\|^2 = \left\| \sum_{j=1}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \right\|^2$$

$$\|\hat{\mathbf{w}} - \hat{\mathbf{w}}_\delta\|^2 = \left\| \sum_{j=1}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \right\|^2$$

$$= \sigma_1^{-2} \left| u^{(1)}, y - y_\delta \right|^2 + \sigma_2^{-2} \left| u^{(2)}, y - y_\delta \right|^2 + \left\| \sum_{j=3}^{d+1} \sigma_j^{-1} v^{(j)} \langle u^{(j)}, y - y_\delta \rangle \right\|^2$$

$$\|\hat{\mathbf{w}} - \hat{\mathbf{w}}_\delta\|^2 = \left\| \sum_{j=1}^{d+1} \sigma_j^{-1} \mathbf{v}^{(j)} \langle \mathbf{u}^{(j)}, (\mathbf{y} - \mathbf{y}_\delta) \rangle \right\|^2$$

$$= \sigma_1^{-2} \left| u^{(1)}, y - y_\delta \right|^2 + \sigma_2^{-2} \left| u^{(2)}, y - y_\delta \right|^2 + \left\| \sum_{j=3}^{d+1} \sigma_j^{-1} v^{(j)} \langle u^{(j)}, y - y_\delta \rangle \right\|^2$$

$$= \sum_{j=1}^{d+1} \sigma_j^{-2} \left| u^{(j)}, y - y_\delta \right|^2$$

Problem 4

Problem 4. Set up a linear regression problem of the form

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathbb{R}^2} \left\{ \frac{1}{2s} \sum_{i=1}^{2} |w^{(0)} + w^{(1)} x^{(i)} - y^{(i)}|^2 \right\}, \tag{3}$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = 1 - c$ and $y^{(1)} = 1$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 1 + c$ and $y^{(2)} = 1$ for some constant $c > 0$.

1. Derive the normal equation for this problem.
2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.
3. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.
4. Repeat the previous exercise, but this time assume you make an error in your measurement. Consider two cases of the new, perturbed measurements
   - $\mathbf{y}_\delta$ reads $y_\delta^{(1)} = 1 - \varepsilon$, $y_\delta^{(2)} = 1 + \varepsilon$.
   - $\mathbf{y}_\delta$ reads $y_\delta^{(1)} = 1 + \varepsilon$, $y_\delta^{(2)} = 1 + \varepsilon$.
5. In both cases compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm and compare with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?
6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

   Hint: make a use of the SVD and use singular vectors you have obtained earlier.

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = 1 - c$ and $y^{(1)} = 1$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 1 + c$ and $y^{(2)} = 1$ for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X = \begin{pmatrix} 1 & 1-c \\ 1 & 1+c \end{pmatrix} \qquad X^\top = \begin{pmatrix} 1 & 1 \\ 1-c & 1+c \end{pmatrix} \qquad y = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = 1 - c$ and $y^{(1)} = 1$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 1 + c$ and $y^{(2)} = 1$ for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X = \begin{pmatrix} 1 & 1-c \\ 1 & 1+c \end{pmatrix} \qquad X^\top = \begin{pmatrix} 1 & 1 \\ 1-c & 1+c \end{pmatrix} \qquad y = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$X^\top X \hat{w} = \begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = X^\top y = \begin{pmatrix} 1 & 1 \\ 1-c & 1+c \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

for data points $(x^{(1)}, y^{(1)})$ with $x^{(1)} = 1 - c$ and $y^{(1)} = 1$, $(x^{(2)}, y^{(2)})$ with $x^{(2)} = 1 + c$ and $y^{(2)} = 1$ for some constant $c > 0$.

1. Derive the normal equation for this problem.

$$X = \begin{pmatrix} 1 & 1-c \\ 1 & 1+c \end{pmatrix} \qquad X^\top = \begin{pmatrix} 1 & 1 \\ 1-c & 1+c \end{pmatrix} \qquad y = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$X^\top X \hat{w} = \begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = X^\top y = \begin{pmatrix} 1 & 1 \\ 1-c & 1+c \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$X^\top X v_i = \sigma_i^2 v_i$$

2. For the matrix **X** you have set up find its singular values and left-/right- singular vectors.

$$X^\top X v_i = \sigma_i^2 v_i$$

$$\det[X^\top X - \sigma_i^2 I] = 0$$

2. For the matrix **X** you have set up find its singular values and left-/right- singular vectors.

$$X^\top X v_i = \sigma_i^2 v_i$$

$$\det[X^\top X - \sigma_i^2 I] = 0$$

$$\det \begin{vmatrix} 2-\sigma_i^2 & 2 \\ 2 & 2+2c^2-\sigma_i^2 \end{vmatrix} = 0$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$X^\top X v_i = \sigma_i^2 v_i$$

$$\det[X^\top X - \sigma_i^2 I] = 0$$

$$\det \begin{vmatrix} 2 - \sigma_i^2 & 2 \\ 2 & 2 + 2c^2 - \sigma_i^2 \end{vmatrix} = 0$$

$$\sigma_i^4 - 2(2 + c^2)\sigma_i^2 + 4c^2 = 0$$

2. For the matrix **X** you have set up find its singular values and left-/right- singular vectors.

$$X^\top X v_i = \sigma_i^2 v_i$$

$$\det[X^\top X - \sigma_i^2 I] = 0$$

$$\det \begin{vmatrix} 2-\sigma_i^2 & 2 \\ 2 & 2+2c^2-\sigma_i^2 \end{vmatrix} = 0$$

$$\sigma_i^4 - 2(2+c^2)\sigma_i^2 + 4c^2 = 0$$

$$\begin{cases} \sigma_1 = \sqrt{c^2+2+\sqrt{c^4+4}}, \\ \sigma_2 = \sqrt{c^2+2-\sqrt{c^4+4}} \end{cases}$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$X^\top X v_i = \sigma_i^2 v_i$$

2. For the matrix **X** you have set up find its singular values and left-/right- singular vectors.

$$X^\top X v_i = \sigma_i^2 v_i$$

$$\begin{pmatrix} 2 & 2 \\ 2 & 2 + 2c^2 \end{pmatrix} \begin{pmatrix} v_1^{(j)} \\ v_2^{(j)} \end{pmatrix} = \sigma_j^2 \begin{pmatrix} v_1^{(j)} \\ v_2^{(j)} \end{pmatrix}$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$X^\top X v_i = \sigma_i^2 v_i$$

$$\begin{pmatrix} 2 & 2 \\ 2 & 2 + 2c^2 \end{pmatrix} \begin{pmatrix} v_1^{(j)} \\ v_2^{(j)} \end{pmatrix} = \sigma_j^2 \begin{pmatrix} v_1^{(j)} \\ v_2^{(j)} \end{pmatrix}$$

$$2v_1^{(j)} + 2v_2^{(j)} = \sigma_j^2 v_1^{(j)}$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$X^\top X v_i = \sigma_i^2 v_i$$

$$\begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \begin{pmatrix} v_1^{(j)} \\ v_2^{(j)} \end{pmatrix} = \sigma_j^2 \begin{pmatrix} v_1^{(j)} \\ v_2^{(j)} \end{pmatrix}$$

$$2v_1^{(j)} + 2v_2^{(j)} = \sigma_j^2 v_1^{(j)}$$

$$v_2^{(j)} = \frac{\sigma_j^2 - 2}{2} v_1^{(j)}$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

2. For the matrix **X** you have set up find its singular values and left-/right- singular vectors.

$$v_2^{(j)} = \frac{\sigma_j^2 - 2}{2} v_1^{(j)}$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$v_2^{(j)} = \frac{\sigma_j^2 - 2}{2} v_1^{(j)}$$

$$v^{(j)} = \left(\gamma, \frac{\sigma_j^2 - 2}{2} \gamma \right)^{\top}$$

2. For the matrix **X** you have set up find its singular values and left-/right- singular vectors.

$$v_2^{(j)} = \frac{\sigma_j^2 - 2}{2} v_1^{(j)}$$

$$v^{(j)} = \left(\gamma, \frac{\sigma_j^2 - 2}{2} \gamma \right)^\top$$

$$\| v^{(j)} \|^2 = 1$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$v_2^{(j)} = \frac{\sigma_j^2 - 2}{2} v_1^{(j)}$$

$$v^{(j)} = \left(\gamma, \frac{\sigma_j^2 - 2}{2} \gamma \right)^{\top}$$

$$\|v^{(j)}\|^2 = 1$$

$$\|v^{(j)}\|^2 = \gamma^2 + \frac{(\sigma_j - 2)^2}{4} \gamma^2 = 1$$

2. For the matrix **X** you have set up find its singular values and left-/right- singular vectors.

$$v_2^{(j)} = \frac{\sigma_j^2 - 2}{2} v_1^{(j)}$$

$$v^{(j)} = \left(\gamma, \frac{\sigma_j^2 - 2}{2} \gamma \right)^\top$$

$$\|v^{(j)}\|^2 = 1$$

$$\|v^{(j)}\|^2 = \gamma^2 + \frac{(\sigma_j - 2)^2}{4} \gamma^2 = 1 \qquad \to \gamma^2 = \frac{4}{4 + (\sigma_j^2 - 2)^2}$$

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

2. For the matrix $\mathbf{X}$ you have set up find its singular values and left-/right- singular vectors.

$$\gamma^2 = \frac{4}{4 + (\sigma_j^2 - 2)^2}$$

2. For the matrix **X** you have set up find its singular values and left-/right- singular vectors.

$$\gamma^2 = \frac{4}{4 + (\sigma_j^2 - 2)^2}$$

$$v^{(j)} = \left(\gamma, \frac{\sigma_j^2 - 2}{2}\gamma\right)^\top$$

2. For the matrix **X** you have set up find its singular values and left-/right- singular vectors.

$$\gamma^2 = \frac{4}{4 + (\sigma_j^2 - 2)^2}$$

$$v^{(j)} = \left(\gamma, \frac{\sigma_j^2 - 2}{2}\gamma\right)^{\mathsf{T}}$$

$$v^{(j)} = \left(\frac{2}{\sqrt{4 + (\sigma_j^2 - 2)^2}}, \frac{\sigma_j^2 - 2}{\sqrt{4 + (\sigma_j^2 - 2)^2}}\right)^{\mathsf{T}}$$

2. For the matrix **X** you have set up find its singular values and left-/right- singular vectors.

2. For the matrix **X** you have set up find its singular values and left-/right- singular vectors.

$$u^{(j)} = \sigma_j^{-1} X v^{(j)}$$

2. For the matrix **X** you have set up find its singular values and left-/right- singular vectors.

$$u^{(j)} = \sigma_j^{-1} X v^{(j)}$$

$$\begin{pmatrix} u_1^{(j)} \\ u_2^{(j)} \end{pmatrix} = \sigma_j^{-1} \begin{pmatrix} 1 & 1-c \\ 1 & 1+c \end{pmatrix} \begin{pmatrix} \dfrac{2}{\sqrt{4+(\sigma_j^2-2)^2}} \\ \dfrac{\sigma_j^2-2}{\sqrt{4+(\sigma_j^2-2)^2}} \end{pmatrix}$$

2. For the matrix **X** you have set up find its singular values and left-/right- singular vectors.

$$u^{(j)} = \sigma_j^{-1} X v^{(j)}$$

$$\begin{pmatrix} u_1^{(j)} \\ u_2^{(j)} \end{pmatrix} = \sigma_j^{-1} \begin{pmatrix} 1 & 1-c \\ 1 & 1+c \end{pmatrix} \begin{pmatrix} \dfrac{2}{\sqrt{4+(\sigma_j^2-2)^2}} \\ \dfrac{\sigma_j^2-2}{\sqrt{4+(\sigma_j^2-2)^2}} \end{pmatrix}$$

$$\begin{pmatrix} u_1^{(j)} \\ u_2^{(j)} \end{pmatrix} = \sigma_j^{-1} \begin{pmatrix} \dfrac{\sigma_j^2(1-c)+2c}{\sqrt{4+(\sigma_j^2-2)^2}} \\ \dfrac{\sigma_j^2(1+c)-2c}{\sqrt{4+(\sigma_j^2-2)^2}} \end{pmatrix}$$

3. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.

3. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.

$$\begin{pmatrix} 2 & 2 \\ 2 & 2 + 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

3. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.

$$\begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

$$w_0 + w_1 = 1$$

3. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.

$$\begin{pmatrix} 2 & 2 \\ 2 & 2 + 2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

$$w_0 + w_1 = 1$$

$$w_0 + (1 + c^2)w_1 = 1$$

3. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.

$$\begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

$$w_0 + w_1 = 1 \qquad\qquad w_0 = 1 - w_1$$

$$w_0 + (1 + c^2)w_1 = 1$$

3. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.

$$\begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

$$w_0 + w_1 = 1 \qquad w_0 = 1 - w_1$$

$$w_0 + (1 + c^2)w_1 = 1 \qquad 1 - w_1 + (1 + c^2)w_1 = 1$$

3. Solve the normal equations for your weights $\hat{\mathbf{w}} = (\hat{w}^{(0)}, \hat{w}^{(1)})^\top$.

$$\begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

$$w_0 + w_1 = 1 \qquad w_0 = 1 - w_1$$

$$w_0 + (1 + c^2)w_1 = 1 \qquad 1 - w_1 + (1 + c^2)w_1 = 1$$

$$\hat{w} = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

4. Repeat the previous exercise, but this time assume you make an error in your measurement. Consider two cases of the new, perturbed measurements

- $\mathbf{y}_\delta$ reads $y_\delta^{(1)} = 1 - \varepsilon$, $y_\delta^{(2)} = 1 + \varepsilon$.
- $\mathbf{y}_\delta$ reads $y_\delta^{(1)} = 1 + \varepsilon$, $y_\delta^{(2)} = 1 + \varepsilon$.

4. Repeat the previous exercise, but this time assume you make an error in your measurement. Consider two cases of the new, perturbed measurements

- $\mathbf{y}_\delta$ reads $y_\delta^{(1)} = 1 - \varepsilon$, $y_\delta^{(2)} = 1 + \varepsilon$.
- $\mathbf{y}_\delta$ reads $y_\delta^{(1)} = 1 + \varepsilon$, $y_\delta^{(2)} = 1 + \varepsilon$.

Repeating the previous exercise with the perturbed data $\mathbf{y}_\delta = \begin{pmatrix} 1-\varepsilon & 1+\varepsilon \end{pmatrix}^\top$ yields the normal equation

$$\begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \hat{\mathbf{w}}_\delta = \begin{pmatrix} 1 & 1 \\ 1-c & 1+c \end{pmatrix} \mathbf{y}_\delta = \begin{pmatrix} 2 \\ 2+2c\varepsilon \end{pmatrix},$$

with the solution

$$\hat{\mathbf{w}}_\delta = \begin{pmatrix} 1 - \frac{\varepsilon}{c} \\ \frac{\varepsilon}{c} \end{pmatrix}.$$

4. Repeat the previous exercise, but this time assume you make an error in your measurement. Consider two cases of the new, perturbed measurements

- $\mathbf{y}_\delta$ reads $y_\delta^{(1)} = 1 - \varepsilon$, $y_\delta^{(2)} = 1 + \varepsilon$.
- $\mathbf{y}_\delta$ reads $y_\delta^{(1)} = 1 + \varepsilon$, $y_\delta^{(2)} = 1 + \varepsilon$.

4. Repeat the previous exercise, but this time assume you make an error in your measurement. Consider two cases of the new, perturbed measurements

- $\mathbf{y}_\delta$ reads $y_\delta^{(1)} = 1 - \varepsilon$, $y_\delta^{(2)} = 1 + \varepsilon$.
- $\mathbf{y}_\delta$ reads $y_\delta^{(1)} = 1 + \varepsilon$, $y_\delta^{(2)} = 1 + \varepsilon$.

For the perturbed data $\mathbf{y}_\delta = \begin{pmatrix} 1+\varepsilon & 1+\varepsilon \end{pmatrix}^\top$ the normal equation takes the form

$$\begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \hat{\mathbf{w}}_\delta = \begin{pmatrix} 1 & 1 \\ 1-c & 1+c \end{pmatrix} \mathbf{y}_\delta$$
$$= \begin{pmatrix} 2+2\varepsilon \\ 2+2\varepsilon \end{pmatrix},$$

with the solution

$$\hat{\mathbf{w}}_\delta = \begin{pmatrix} 1+\varepsilon \\ 0 \end{pmatrix}.$$

5. In both cases compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm and compare with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?

5. In both cases compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm and compare with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?

$$\hat{w} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \hat{w}_\delta = \begin{pmatrix} 1 - \frac{\epsilon}{c} \\ \frac{\epsilon}{c} \end{pmatrix}$$

5. In both cases compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm and compare with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?

$$\hat{w} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \hat{w}_\delta = \begin{pmatrix} 1 - \frac{\epsilon}{c} \\ \frac{\epsilon}{c} \end{pmatrix} \qquad \|\hat{w} - \hat{w}_\delta\| = \sqrt{\frac{\epsilon^2}{c^2} + \frac{\epsilon^2}{c^2}} = \frac{\epsilon}{c}\sqrt{2}$$

5. In both cases compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm and compare with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?

$$\hat{w} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \hat{w}_\delta = \begin{pmatrix} 1 - \frac{\epsilon}{c} \\ \frac{\epsilon}{c} \end{pmatrix} \qquad \|\hat{w} - \hat{w}_\delta\| = \sqrt{\frac{\epsilon^2}{c^2} + \frac{\epsilon^2}{c^2}} = \frac{\epsilon}{c}\sqrt{2}$$

$$\hat{w} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \hat{w}_\delta = \begin{pmatrix} 1 + \epsilon \\ 0 \end{pmatrix}$$

5. In both cases compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm and compare with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?

$$\hat{w} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \hat{w}_\delta = \begin{pmatrix} 1 - \frac{\epsilon}{c} \\ \frac{\epsilon}{c} \end{pmatrix} \qquad \|\hat{w} - \hat{w}_\delta\| = \sqrt{\frac{\epsilon^2}{c^2} + \frac{\epsilon^2}{c^2}} = \frac{\epsilon}{c}\sqrt{2}$$

$$\hat{w} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \hat{w}_\delta = \begin{pmatrix} 1 + \epsilon \\ 0 \end{pmatrix} \qquad \|\hat{w} - \hat{w}_\delta\| = \sqrt{\epsilon^2} = \epsilon$$

6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

Hint: make a use of the SVD and use singular vectors you have obtained earlier.

$$\|\hat{\mathbf{w}} - \hat{\mathbf{w}}_\delta\|^2 = \sum_{j=1}^{d+1} \sigma_j^{-2} \left| \langle \mathbf{u}^{(2)}, \mathbf{y} - \mathbf{y}_\delta \rangle \right|^2 .$$

6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

Hint: make a use of the SVD and use singular vectors you have obtained earlier.

$$\|\hat{\mathbf{w}} - \hat{\mathbf{w}}_\delta\|^2 = \sum_{j=1}^{d+1} \sigma_j^{-2} \left|\langle \mathbf{u}^{(2)}, \mathbf{y} - \mathbf{y}_\delta \rangle\right|^2 .$$

The smallest singular value is the most important as well as the scalar product!

6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

   Hint: make a use of the SVD and use singular vectors you have obtained earlier.

6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

Hint: make a use of the SVD and use singular vectors you have obtained earlier.

First case $\quad y = (1,1)^\top \quad y_\delta = (1-\epsilon, 1+\epsilon)^\top$

6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

Hint: make a use of the SVD and use singular vectors you have obtained earlier.

$$First \quad case \qquad y = (1,1)^\top \qquad y_\delta = (1 - \epsilon, 1 + \epsilon)^\top \qquad y - y_\delta = \epsilon(1, -1)^\top$$

6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

Hint: make a use of the SVD and use singular vectors you have obtained earlier.

First case $y = (1,1)^\top$ $y_\delta = (1-\epsilon, 1+\epsilon)^\top$ $y - y_\delta = \epsilon(1, -1)^\top$

Second case $y = (1,1)^\top$ $y_\delta = (1+\epsilon, 1+\epsilon)^\top$

6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

Hint: make a use of the SVD and use singular vectors you have obtained earlier.

First case $y = (1,1)^\top$ $y_\delta = (1-\epsilon, 1+\epsilon)^\top$ $y - y_\delta = \epsilon(1,-1)^\top$

Second case $y = (1,1)^\top$ $y_\delta = (1+\epsilon, 1+\epsilon)^\top$ $y - y_\delta = -\epsilon(1,1)^\top$

6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

Hint: make a use of the SVD and use singular vectors you have obtained earlier.

First case

$$y - y_\delta = \epsilon(1, -1)^\top$$

6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

Hint: make a use of the SVD and use singular vectors you have obtained earlier.

First case $y - y_\delta = \epsilon(1, -1)^\top$

$$u^{(2)} = \sigma_2^{-1} \left(\frac{\sigma_2^2(1-c)+2c}{\sqrt{4+(\sigma_2^2-2)^2}}, \frac{\sigma_2^2(1+c)-2c}{\sqrt{4+(\sigma_2^2-2)^2}} \right)^\top$$

6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

Hint: make a use of the SVD and use singular vectors you have obtained earlier.

First case $\quad y - y_\delta = \epsilon(1, -1)^\top$

$$u^{(2)} = \sigma_2^{-1}\left(\frac{\sigma_2^2(1-c)+2c}{\sqrt{4+(\sigma_2^2-2)^2}}, \frac{\sigma_2^2(1+c)-2c}{\sqrt{4+(\sigma_2^2-2)^2}}\right)^\top$$

$$\langle u^{(2)}, y - y_\delta\rangle = \epsilon\sigma_2^{-1}\left(\frac{\sigma_2^2(1-c)+2c}{\sqrt{4+(\sigma_2^2-2)^2}} - \frac{\sigma_2^2(1+c)-2c}{\sqrt{4+(\sigma_2^2-2)^2}}\right)$$

6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

Hint: make a use of the SVD and use singular vectors you have obtained earlier.

First case

$$\langle u^{(2)}, y - y_\delta \rangle = \epsilon \sigma_2^{-1} \left(\frac{\sigma_2^2(1-c) + 2c}{\sqrt{4 + (\sigma_2^2 - 2)^2}} - \frac{\sigma_2^2(1+c) - 2c}{\sqrt{4 + (\sigma_2^2 - 2)^2}} \right)$$

6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

Hint: make a use of the SVD and use singular vectors you have obtained earlier.

First case

$$\langle u^{(2)}, y - y_\delta \rangle = \epsilon \sigma_2^{-1} \left(\frac{\sigma_2^2(1-c)+2c}{\sqrt{4+(\sigma_2^2-2)^2}} - \frac{\sigma_2^2(1+c)-2c}{\sqrt{4+(\sigma_2^2-2)^2}} \right)$$

$$= \epsilon \sigma_2^{-1} \left(\frac{4c - 2c\sigma_2^2}{\sqrt{4+(\sigma_2^2-2)^2}} \right)$$

6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

Hint: make a use of the SVD and use singular vectors you have obtained earlier.

Second case

$$y - y_\delta = -\epsilon(1,1)^\top$$

6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

Hint: make a use of the SVD and use singular vectors you have obtained earlier.

Second case

$$y - y_\delta = -\epsilon(1,1)^\top$$

$$u^{(2)} = \sigma_2^{-1}\left(\frac{\sigma_2^2(1-c)+2c}{\sqrt{4+(\sigma_2^2-2)^2}}, \frac{\sigma_2^2(1+c)-2c}{\sqrt{4+(\sigma_2^2-2)^2}}\right)^\top$$

6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

Hint: make a use of the SVD and use singular vectors you have obtained earlier.

Second case

$$y - y_\delta = -\epsilon(1,1)^\top$$

$$u^{(2)} = \sigma_2^{-1}\left(\frac{\sigma_2^2(1-c)+2c}{\sqrt{4+(\sigma_2^2-2)^2}}, \frac{\sigma_2^2(1+c)-2c}{\sqrt{4+(\sigma_2^2-2)^2}}\right)^\top$$

$$\langle u^{(2)}, y - y_\delta\rangle = -\epsilon\sigma_2^{-1}\left(\frac{\sigma_2^2(1-c)+2c}{\sqrt{4+(\sigma_2^2-2)^2}} + \frac{\sigma_2^2(1+c)-2c}{\sqrt{4+(\sigma_2^2-2)^2}}\right)$$

6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

Hint: make a use of the SVD and use singular vectors you have obtained earlier.

Second case

6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

Hint: make a use of the SVD and use singular vectors you have obtained earlier.

Second case

$$\langle u^{(2)}, y - y_\delta \rangle = -\epsilon \sigma_2^{-1} \left(\frac{\sigma_2^2(1-c)+2c}{\sqrt{4+(\sigma_2^2-2)^2}} + \frac{\sigma_2^2(1+c)-2c}{\sqrt{4+(\sigma_2^2-2)^2}} \right)$$

6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

Hint: make a use of the SVD and use singular vectors you have obtained earlier.

Second case

$$\langle u^{(2)}, y - y_\delta \rangle = -\epsilon \sigma_2^{-1} \left(\frac{\sigma_2^2(1-c)+2c}{\sqrt{4+(\sigma_2^2-2)^2}} + \frac{\sigma_2^2(1+c)-2c}{\sqrt{4+(\sigma_2^2-2)^2}} \right)$$

$$= -\epsilon \sigma_2^{-1} \left(\frac{2\sigma_2^2}{\sqrt{4+(\sigma_2^2-2)^2}} \right)$$

6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

Hint: make a use of the SVD and use singular vectors you have obtained earlier.

Second case

$$\langle u^{(2)}, y - y_\delta \rangle = -\epsilon\sigma_2^{-1}\left(\frac{\sigma_2^2(1-c)+2c}{\sqrt{4+(\sigma_2^2-2)^2}} + \frac{\sigma_2^2(1+c)-2c}{\sqrt{4+(\sigma_2^2-2)^2}}\right)$$

$$= -\epsilon\sigma_2^{-1}\left(\frac{2\sigma_2^2}{\sqrt{4+(\sigma_2^2-2)^2}}\right)$$

$$= -2\epsilon\left(\frac{2\sigma_2}{\sqrt{4+(\sigma_2^2-2)^2}}\right)$$

6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

Hint: make a use of the SVD and use singular vectors you have obtained earlier.

First case $$\langle u^{(2)}, y - y_\delta \rangle = \epsilon \sigma_2^{-1} \left(\frac{4c - 2c\sigma_2^2}{\sqrt{4 + (\sigma_2^2 - 2)^2}} \right)$$

Second case $$\langle u^{(2)}, y - y_\delta \rangle = = -2\epsilon \left(\frac{2\sigma_2}{\sqrt{4 + (\sigma_2^2 - 2)^2}} \right)$$

6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

Hint: make a use of the SVD and use singular vectors you have obtained earlier.

$$\textit{First case} \quad \langle u^{(2)}, y - y_\delta \rangle = \epsilon \sigma_2^{-1} \left(\frac{4c - 2c\sigma_2^2}{\sqrt{4 + (\sigma_2^2 - 2)^2}} \right) \xrightarrow{\sigma_2 \to 0} \langle u^{(2)}, y - y_\delta \rangle \sim \sigma_2^{-1}$$

$$\textit{Second case} \quad \langle u^{(2)}, y - y_\delta \rangle = = -2\epsilon \left(\frac{2\sigma_2}{\sqrt{4 + (\sigma_2^2 - 2)^2}} \right)$$

6. Explain why do you observe such a huge difference between the two cases when $c \to 0$?

Hint: make a use of the SVD and use singular vectors you have obtained earlier.

$$\textit{First case} \quad \langle u^{(2)}, y - y_\delta \rangle = \epsilon \sigma_2^{-1} \left(\frac{4c - 2c\sigma_2^2}{\sqrt{4 + (\sigma_2^2 - 2)^2}} \right) \xrightarrow{\sigma_2 \to 0} \langle u^{(2)}, y - y_\delta \rangle \sim \sigma_2^{-1}$$

$$\textit{Second case} \quad \langle u^{(2)}, y - y_\delta \rangle = = -2\epsilon \left(\frac{2\sigma_2}{\sqrt{4 + (\sigma_2^2 - 2)^2}} \right) \xrightarrow{\sigma_2 \to 0} \langle u^{(2)}, y - y_\delta \rangle \sim \sigma_2$$