Queen Mary University of London

MTH786, Semester A, 2023/24
Solutions of coursework 3 **N. Perra**

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.

Solutions:

1. 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:

$\forall x \in \mathcal{C}, \quad \lambda \in [0,1]$:

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

Hence, the sum of two convex functions is also convex.

2. 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. We compute

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

which proves the assertion.

4. 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. 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.

6. 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. We proceed in similar fashion as in the previous exercise. We point out that the MSE can be written as $\text{MSE}(w) = g(h(w))$ 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. Set up a linear regression problem of the form

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

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 $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$?

Solutions:

1. The data matrix X for the points specified in the problem description reads

$$X = \begin{pmatrix} 1 & -c \\ 1 & 0 \\ 1 & c \end{pmatrix}. \tag{2}$$

From the lecture notes we know that the normal equation $X^\top X\hat{w} = X^\top y$ solves Problem (1). For X as defined in Equation (4) and $y := \begin{pmatrix} 2 & 2 & 2 \end{pmatrix}^\top$ we then calculate

$$\begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix}\hat{w} = \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} y$$
$$= \begin{pmatrix} 6 \\ 0 \end{pmatrix}.$$

2. We easily solve the previous equation for $\hat{w}$ and obtain

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

hence, $\hat{w}_0 = 2$ and $\hat{w}^{(1)} = 0$. We obtain a line with slope zero and a constant translation of two.

3. Repeating the previous two exercises with the perturbed data $y^\delta = \begin{pmatrix} 2+\varepsilon & 2+\varepsilon & 2-\varepsilon \end{pmatrix}^\top$ yields the normal equation

$$\begin{aligned} \begin{pmatrix} 3 & 0 \\ 0 & 2c^2 \end{pmatrix} \hat{w}_\delta &= \begin{pmatrix} 1 & 1 & 1 \\ -c & 0 & c \end{pmatrix} y^\delta \\ &= \begin{pmatrix} 6+\varepsilon \\ -2\,c\,\varepsilon \end{pmatrix} , \end{aligned}$$

with the solution

$$\hat{w}_\delta = \begin{pmatrix} 2 + \frac{\varepsilon}{3} \\ -\frac{\varepsilon}{c} \end{pmatrix} .$$

4. The error in terms of the Euclidean norm reads

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

5. 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. 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.

Solutions:

1. Based on the singular vector decomposition representations of $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$, we have

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

Here, the last equality follows from the orthonormality of the singular vectors $\mathbf{v}^{(j)}$, which implies $\|\mathbf{v}^{(1)}\|^2 = 1$ and $\langle \mathbf{v}^{(1)}, \mathbf{v}^{(j)} \rangle = 0$ for all $j \neq 1$. Recursively (or by induction) we can repeat the same argument for the squared norm of the remaining sum, and, thus, verify the statement.

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.

Solutions:

1. The data matrix X for the points specified in the problem description reads

$$\mathbf{X} = \begin{pmatrix} 1 & 1-c \\ 1 & 1+c \end{pmatrix}. \tag{4}$$

From the lecture notes we know that the normal equation $\mathbf{X}^\top \mathbf{X} \hat{w} = \mathbf{X}^\top \mathbf{y}$ solves Problem (1). For $\mathbf{X}$ as defined in Equation (4) and $\mathbf{y} := \begin{pmatrix} 1 & 1 \end{pmatrix}^\top$ we then calculate

$$\begin{pmatrix} 2 & 2 \\ 2 & 2+2c^2 \end{pmatrix} \hat{\mathbf{w}} = \begin{pmatrix} 1 & 1 \\ 1-c & 1+c \end{pmatrix} \mathbf{y}$$
$$= \begin{pmatrix} 2 \\ 2 \end{pmatrix}.$$

2. Singular value of matrix $\mathbf{X}$ can be found as eigenvalues of matrix $\mathbf{X}^\top \mathbf{X}$. Solving

$$\det\left(\mathbf{X}^\top \mathbf{X} - \sigma^2 I\right) = 0,$$

one obtains

$$\sigma^4 - \left(4 + 2c^2\right)\sigma^2 + 4c^2 = 0.$$

Solutions of the above are

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

The right singular vectors of matrix $\mathbf{X}$ are eigenvectors of $\mathbf{X}^\top \mathbf{X}$. These can be found by solving

$$\mathbf{X}^\top \mathbf{X} \mathbf{v}^{(j)} = \sigma_j^2 \mathbf{v}^{(j)} \Leftrightarrow \left(2 - \sigma_j^2\right)\mathbf{v}_1^{(j)} + 2\mathbf{v}_2^{(j)} = 0 \Rightarrow \mathbf{v}_2^{(j)} = \frac{\sigma_j^2 - 2}{2}\mathbf{v}_1^{(j)}$$

$$\mathbf{v}^{(j)} = \begin{pmatrix} \dfrac{2}{\sqrt{4 + \left(\sigma_j^2 - 2\right)^2}} & \dfrac{\sigma_j^2 - 2}{\sqrt{4 + \left(\sigma_j^2 - 2\right)^2}} \end{pmatrix}^\top.$$

For the left singular vectors we first need to calculate the product $\mathbf{X}\mathbf{X}^\top$ and then find corresponding eigenvectors. It is easy to check that

$$\mathbf{X}\mathbf{X}^\top = \begin{pmatrix} 2 - 2c + c^2 & 2 - c^2 \\ 2 - c^2 & 2 + 2c + c^2 \end{pmatrix}$$

and the left singular vectors then solve

$$\left(2-2c+c^2-\sigma_j^2\right)\mathbf{u}_1^{(j)}+\left(2-c^2\right)\mathbf{u}_2^{(j)}=0 \Rightarrow \mathbf{u}_2^{(j)}=\frac{\sigma_j^2-2-c^2+2c}{2-c^2}\mathbf{u}_1^{(j)},$$

$$\mathbf{u}^{(j)}=\left(\frac{2-c^2}{\sqrt{(2-c^2)^2+\left(\sigma_j^2-c^2-2+2c\right)^2}} \quad \frac{\sigma_j^2-c^2-2+2c}{\sqrt{(2-c^2)^2+\left(\sigma_j^2-c^2-2+2c\right)^2}}\right)^\top .$$

Remark: the expressions are quite nasty, but we will work with these in small c case only. If c is small, then $\sigma_1^2 \approx 4+c^2$, $\sigma_2^2 \approx c^2$ and

$$\mathbf{v}^{(1)} \approx \left(\frac{1}{\sqrt{2}} \quad \frac{1}{\sqrt{2}}\right)^\top, \quad \mathbf{v}^{(2)} \approx \left(\frac{1}{\sqrt{2}} \quad -\frac{1}{\sqrt{2}}\right)^\top .$$

$$\mathbf{u}^{(1)} \approx \left(\frac{1}{\sqrt{2}} \quad \frac{1}{\sqrt{2}}\right)^\top, \quad \mathbf{u}^{(2)} \approx \left(\frac{1}{\sqrt{2}} \quad -\frac{1}{\sqrt{2}}\right)^\top .$$

3. We easily solve the normal equation for $\hat{\mathbf{w}}$ and obtain

$$\hat{\mathbf{w}}=\begin{pmatrix}1\\0\end{pmatrix};$$

hence, $\hat{w}_0=1$ and $\hat{w}_1=0$. We obtain a line with slope zero and a constant translation of two.

4.
   - Repeating the previous exercise with the perturbed data $\mathbf{y}_\delta=\left(1-\varepsilon \quad 1+\varepsilon\right)^\top$ yields the normal equation

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

   with the solution

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

   - For the perturbed data $\mathbf{y}_\delta=\left(1+\varepsilon \quad 1+\varepsilon\right)^\top$ 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. • The error in terms of the Euclidean norm reads

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

• In the second case it reads

$$\|\hat{\mathbf{w}} - \hat{\mathbf{w}}_\delta\| = \sqrt{\varepsilon^2 + 0^2} = \varepsilon.$$

The data error in both cases is equal to $\sqrt{2}\varepsilon$. One can see that in the second case the error in $\hat{\mathbf{w}}$ is just the data error divided by $\sqrt{2}$. So it is of the same magnitude. While in the first case we can have a much higher error if c is small enough.

6. As we have seen in the first exercise, the error can be written as

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

The term that can bring this error to a high value corresponds to the lowest singular value, i.e. σ_2 in our case. To make this term large one should have large $\left\langle \mathbf{u}^{(2)}, (\mathbf{y} - \mathbf{y}_\delta) \right\rangle$. Thus the error in $\hat{\mathbf{w}}$ is bigger if the data perturbation is parallel to $\mathbf{u}^{(2)}$. Now one can check that in the first case we have a perturbation $\mathbf{y} - \mathbf{y}_\delta = \varepsilon \begin{pmatrix} -1 & 1 \end{pmatrix}^\top$ that is parallel to $\mathbf{u}^{(2)}$, while in the second case one has a perturbation $\mathbf{y} - \mathbf{y}_\delta = \varepsilon \begin{pmatrix} 1 & 1 \end{pmatrix}^\top$ that is just orthogonal to $\mathbf{u}^{(2)}$.