MTH786, Semester A, 2023/24

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.

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

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.

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

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

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.