Queen Mary University of London

MTH786, Semester A, 2023/24

Coursework 2 **N. Perra**

Problem 1. In the lecture we discussed different types of error measures. In this task you are asked to compare the stability/robustness of both measures to an outlier. The data samples given are $\left(x^{(1)}, y^{(1)}\right) = (-2, 1)$, $\left(x^{(2)}, y^{(2)}\right) = (-1, 2)$, $\left(x^{(3)}, y^{(3)}\right) = (0, 3)$, $\left(x^{(4)}, y^{(4)}\right) = (1, 4)$.

1. Compute the MSE for the 1-parameter model by hand:

$$\mathrm{MSE}(w^{(0)}) = \frac{1}{2s} \sum_{i=1}^{s} |y^{(i)} - w^{(0)}|^2,$$

for $w^{(0)} \in \{1, 2, 3, 4, 5, 6, 7\}$. Between the above values find $w^{(0)}$ that minimises the MSE. A new data sample $\left(x^{(5)}, y^{(5)}\right) = (2, 20)$ is added. Evaluate new error measure and corresponding minimiser.

You may find it useful to fill in the missing entries of the following table:

	$w^{(0)} = 1$	$w^{(0)} = 2$	$w^{(0)} = 3$	$w^{(0)} = 4$	$w^{(0)} = 5$	$w^{(0)} = 6$	$w^{(0)} = 7$
$y^{(1)} = 1$							
$y^{(2)} = 2$							
$y^{(3)} = 3$							
$y^{(4)} = 4$							
$\mathrm{MSE}(w) \cdot 2s$							
$y^{(5)} = 20$							
$\mathrm{MSE}(w) \cdot 2s$							

Some help: $19^2 = 361, 18^2 = 324, 17^2 = 289, 16^2 = 256, 15^2 = 225, 14^2 = 196, 13^2 = 169$.

2. Repeat the same exercise for what is known as the Mean Absolute Error (MAE), i.e.

$$\mathrm{MAE}(w^{(0)}) = \frac{1}{s} \sum_{i=1}^{s} |y^{(i)} - w^{(0)}| \,.$$

What do you observe, in particular with regards to the outlier $y^{(5)}$?

Problem 2. Assume we are given s i.i.d. samples $x_1, \ldots, x_s$, and we know that they are drawn from a normal distribution with mean μ and variance σ^2. We do not know these two parameters and want to estimate them from the data using the maximum likelihood principle.

1. Write down the likelihood for this data, i.e., the joint probability distribution function $\rho(x_1, \ldots, x_s \,|\, \mu, \sigma^2)$, where the notation reminds us that this PDF depends on the two parameters μ and σ^2.

2. Use the maximum likelihood principle to estimate the parameter μ. More precisely, compute the gradient of the negative log-likelihood with respect to μ, set it to zero and solve for μ. This gives us an estimator $\hat{\mu}$ of μ that depends on the data.

3. Use the maximum likelihood principle to estimate the parameter σ^2. Proceed in the same manner as in section 2, but this time with the parameter σ^2 instead of μ.

4. Verify that $-\nabla \log(\rho(w)) = 0$ automatically implies $\nabla \rho(w) = 0$, regardless of the choice of probability density function ρ.

Problem 3. In the lecture we have seen that the the general MSE cost function for the linear regression is of the form

$$\mathrm{MSE}(\mathbf{W}) = \frac{1}{2s} \|\mathbf{XW} - \mathbf{Y}\|^2,$$

where

$$\mathbf{X} = \begin{pmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \dots & x_d^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \dots & x_d^{(2)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_1^{(s)} & x_2^{(s)} & \dots & x_d^{(s)} \end{pmatrix}, \quad \mathbf{W} = \begin{pmatrix} w_1^{(0)} & w_2^{(0)} & w_3^{(0)} & \dots & w_n^{(0)} \\ w_1^{(1)} & w_2^{(1)} & w_3^{(1)} & \dots & w_n^{(1)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ w_1^{(d)} & w_2^{(d)} & w_3^{(d)} & \dots & w_n^{(d)} \end{pmatrix},$$

$$\mathbf{Y} = \begin{pmatrix} y_1^{(1)} & y_2^{(1)} & \dots & y_n^{(1)} \\ y_1^{(2)} & y_2^{(2)} & \dots & y_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ y_1^{(s)} & y_2^{(s)} & \dots & y_n^{(s)} \end{pmatrix},$$

and the norm $\|\cdot\|$ is a Frobenius norm defined by

$$\|\mathbf{M}\|^2 = \sum_{i,j} m_{i,j}^2.$$

Prove that the gradient of MSE is given by

$$\nabla \mathrm{MSE}(w) = \frac{1}{s} \mathbf{X}^\top (\mathbf{XW} - \mathbf{Y}).$$

Problem 4. Compute the solution of the polynomial regression problem

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

by hand, for the data samples $(x^{(1)}, y^{(1)}) = (0, 0)$, $(x^{(2)}, y^{(2)}) = (1/4, 1)$, $(x^{(3)}, y^{(3)}) = (1/2, 0)$, $(x^{(4)}, y^{(4)}) = (3/4, -1)$ and $(x^{(5)}, y^{(5)}) = (1, 0)$ and choices

1. $d = 1$,

2. $d = 2$,

3. $d = 3$.