Exercise sheet: Linear regression

The following exercises have different levels of difficulty indicated by (*), (**), (***). An exercise with (*) is a simple exercise requiring less time to solve compared to an exercise with (***), which is a more complex exercise.

1. (*) Given the two vectors,

$$\mathbf{x} = \begin{bmatrix} 1.3 \\ -2.0 \\ 4.1 \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} 0.4 \\ -0.8 \\ -1.1 \end{bmatrix}.$$

compute their inner product and their outer product.

2. (**) Let us define a matrix $\mathbf{W}$ of dimensions $n \times m$, a vector $\mathbf{x}$ of dimensions $m \times 1$ and a vector $\mathbf{y}$ of dimensions $n \times 1$. Write the following expression in matrix form

$$\sum_{i=1}^{n}\sum_{j=1}^{m} w_{i,j}x_j + \sum_{j=1}^{m}\sum_{i=1}^{n} y_i w_{i,j}.$$

[HINT: if necessary define a vector of ones $\mathbf{1}_p = [1 \cdots 1]^\top$ of dimensions $p \times 1$, where p can be any number].

3. (***) Show that using the ML criterion, the optimal value for σ_*^2 is given as in slide 40 of Lecture 4, this is,

$$\sigma_*^2 = \frac{1}{N}(\mathbf{y} - \mathbf{X}\mathbf{w}_*)^\top(\mathbf{y} - \mathbf{X}\mathbf{w}_*).$$

4. (*) You are given a dataset with the following instances, $(x_1, y_1) = (0.8, -1.2), (x_2, y_2) = (-0.3, -0.6)$, and $(x_3, y_3) = (0.1, 2.4)$. Find the optimal value $\mathbf{w}_*$ used in ridge regression with a regularisation parameter $\lambda = 0.1$.

5. (***) Consider a regression problem for which each observed output y_n has an associated weight factor $r_n > 0$, such that the mean of weighted squared errors is given as

$$E(\mathbf{w}) = \frac{1}{N}\sum_{n=1}^{N} r_n(y_n - \mathbf{w}^\top\mathbf{x}_n)^2,$$

where $\mathbf{w} = [w_0, \ldots, w_D]^\top$ is the vector of parameters, and $\mathbf{x}_n \in \mathbb{R}^{D+1\times 1}$ with $x_{n,0} = 1$.

   (a) Starting with the expression above, write the mean of weighted squared errors in matrix form. You should include each of the steps necessary to get the matrix form solution. [HINT: a diagonal matrix is a matrix that is zero everywhere except for the entries on its main diagonal. The weight factors $r_n > 0$ can be written as the elements of a diagonal matrix $\mathbf{R}$ of size $N \times N$].

(b) Find the optimal value of $\mathbf{w}$, $\mathbf{w}_*$, that minimises the mean of weighted squared errors. The solution should be in matrix form. Use matrix derivatives.

6. (*) A dataset is used to train a linear regression model with polynomial basis functions $\{\phi_i(x) = x^i\}_{i=1}^{M}$, where $M = 4$. Assume that the weight vector after training is equal to $\mathbf{w}_* = [0.5, -0.8, 1.2, 1.3, -0.3]^\top$. What would be the predicted value for this linear model when the input is $x = 2.5$?

7. (***) Show that the optimal solution for $\mathbf{w}_*$ in ridge regression is given as in slide 63 of Lecture 4, this is,

$$\mathbf{w}_* = \left(\mathbf{X}^\top\mathbf{X} + \frac{\lambda N}{2}\mathbf{I}\right)^{-1}\mathbf{X}^\top\mathbf{y}.$$