The Mathematics of Machine Learning
Homework Set 2

Due 6 March 2024 before 13:00
via Canvas

You are allowed to work on this homework in pairs. One person per pair submits the answers via Canvas. Make sure to put both names on the submission.

1 Theory Exercises

1. [4 pt] Consider linear regression. For each of the following data sets, determine whether the least squares solution

$$\arg\min_{\beta_0,\beta} \quad \sum_{i=1}^{N}(Y_i - X_i^\top \beta - \beta_0)^2$$

is unique:

   (a) $T = (X_1, Y_1), \ldots, (X_3, Y_3) = \left(\begin{pmatrix}2\\3\end{pmatrix}, 5\right), \left(\begin{pmatrix}3\\4\end{pmatrix}, 7\right), \left(\begin{pmatrix}4\\5\end{pmatrix}, 8\right)$

   (b) $T = (X_1, Y_1), \ldots, (X_3, Y_3) = \left(\begin{pmatrix}2\\3\end{pmatrix}, -1\right), \left(\begin{pmatrix}2\\4\end{pmatrix}, -2\right), \left(\begin{pmatrix}2\\5\end{pmatrix}, -3\right)$

2. [4 pt] A common step to prepare data before applying any machine learning method is to center the data points around their mean:

$$X_i' = X_i - \bar{x}, \qquad Y_i' = Y_i - \bar{y},$$

where $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} X_i$ is the mean of the feature vectors, and $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$ is the mean of the response vectors in the training data.

In preparation for lecture 4, we will consider estimators of the form

$$(\hat{\beta}_0, \hat{\beta}) = \arg\min_{(\beta_0,\beta)} \Big(\sum_{i=1}^{N}(Y_i - X_i^\top \beta - \beta_0)^2 + \lambda \mathrm{pen}(\beta) \Big), \tag{1}$$

where $\lambda \geq 0$ is a fixed number and pen is a function of β that does not depend on β_0. (Note that we recover least squares for $\lambda = 0$.) Now consider the following alternative criterion on the centered data, in which we only solve for β:

$$\hat{\beta} = \arg\min_{\beta} \Big(\sum_{i=1}^{N} (Y_i' - (X_i')^\top \beta)^2 + \lambda \mathrm{pen}(\beta) \Big). \tag{2}$$

Show that (2) gives the same solution for $\hat{\beta}$ as (1).
Hint 1: Note that we can write the minimization problem in (1) *as*

$$\min_{\beta} \min_{\beta_0} \Big(\sum_{i=1}^{N} (Y_i - X_i^\top \beta - \beta_0)^2 + \lambda \mathrm{pen}(\beta) \Big).$$

Hint 2: Solve the inner minimization over β_0 for any fixed β and plug the solution back in to eliminate β_0. Show that the remaining optimization over β is equivalent to (2).