Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed Computing

HS 2023

Prof. R. Wattenhofer
Benjamin Estermann

Computational Thinking
Exercise 9

1 Linear Regression

Here is a dataset D with 3 samples. You want to fit a linear model of the form $\hat{f}(x) = w_0 + w_1 x$.

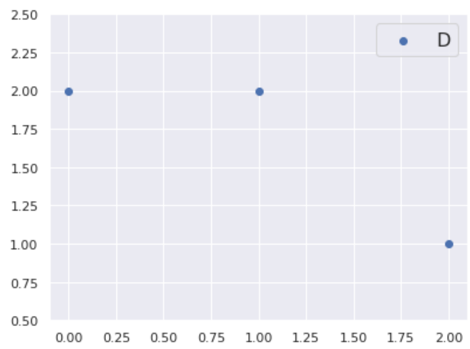

Figure 1: A dataset with 3 samples.

a) Give the weights that minimize the squared error loss and compute the total absolute error and total squared error for them.

b) Can you minimize the absolute error loss? What is the resulting total absolute and total squared error?

2 Polynomial Regression

You are given the following function f (that is only defined on the interval depicted in the figure) and a sample of datapoints D (sampling was done with some noise).

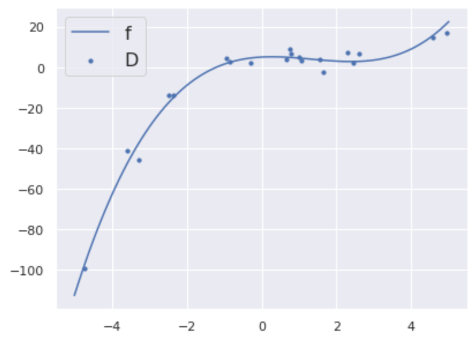

Figure 2: A function and some training data D.

Which one of the following models will result in the lowest bias? And which one in the lowest variance?

a) $\hat{f} = 3$

b) $\hat{f} = w_0$

c) $\hat{f} = w_0 + w_1 x$

d) $\hat{f} = w_0 + w_1 x + w_2 x^2$

e) $\hat{f} = w_0 + w_1 x + w_2 x^2 + w_3 x^3$

3 Ridge Regression

In the lecture we saw that linear regression without regularization has a closed form solution:

$$\boldsymbol{w}^* = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\boldsymbol{y}.$$

Recall from the lecture that Ridge regression minimizes:

$$\min_{\boldsymbol{w}} \left\{ \frac{1}{n} \sum_{(\boldsymbol{x},y)\in D} (y - \boldsymbol{w}^T\boldsymbol{x})^2 + \lambda \sum_{i=0}^{d-1} w_i^2 \right\}$$

a) Show that Ridge regression has the following closed-form solution by differentiating the loss function.

$$\boldsymbol{w}^*_{ridge} = \left(\mathbf{X}^T\mathbf{X} + \lambda n \mathbf{I}\right)^{-1}\mathbf{X}^T\boldsymbol{y}$$

where $\mathbf{I}$ is the $d \times d$ identity matrix.

b) What happens to the weights $\boldsymbol{w}^*_{ridge}$ in the limit as $\lambda \to \infty$?

c) What happens to the weights $\boldsymbol{w}^*_{ridge}$ in the limit as $\lambda \to 0$?

4 Rescaling

Suppose we have a dataset D with 1000 samples and 100 features $\{x_1, x_2, \ldots, x_{100}\}$. Now, we rescale one of these features by multiplying with 10 (say that feature is x_1).

a) Show that the ordinary least-square (OLS) weights remain unchanged for $i > 1$, and that $w_1^{*\prime} = \frac{1}{10} w_1^*$

b) Conclude that the OLS predictions do not change.

c) What about Lasso and Ridge regression? Do the weights change? Do the predictions change?