Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed Computing

HS 2020

Prof. R. Wattenhofer
B. Egressy

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

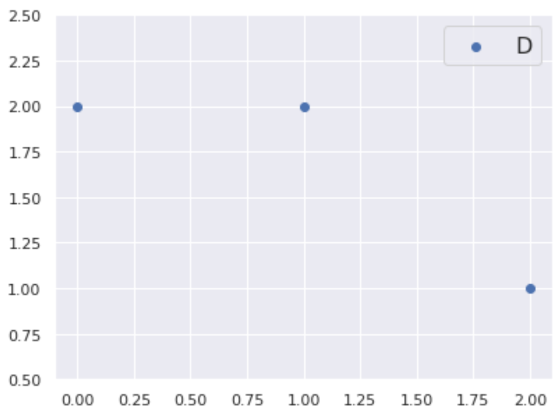

Figure 1: A dataset with 3 samples.

a) Which weights minimize the squared error loss? What is the total absolute error? What is the total squared error?

b) Which weights give a lower absolute error loss? What is the total absolute error? What is the total squared error?

2 Polynomial Regression

Here is a function. You have sampled a set of training data points with noise as shown by the dots.

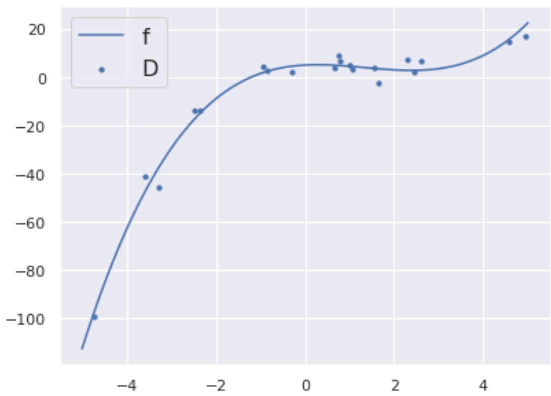

Figure 2: A function and some training data D.

Which model will give you the lowest bias:

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

Which model will give you 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) By differentiating the loss function, show that Ridge regression has the following closed-form solution:

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

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 feature by multiplying with 10 (say that feature is x_1).

a) Show that the 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 with Lasso and Ridge regression? Do the weights change? Do the predictions change?