

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 $\hat{f}(x) = w_0 + w_1x$.

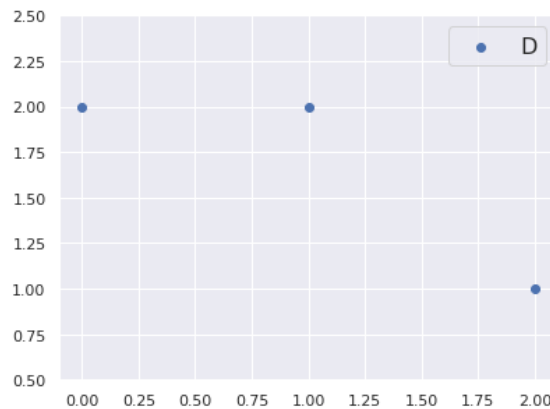


Figure 1: A dataset with 3 samples.

- Which weights minimize the squared error loss? What is the total absolute error? What is the total squared error?
- 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.

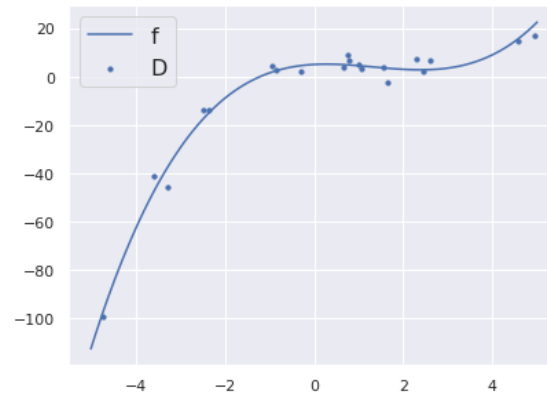


Figure 2: A function and some training data D .

Which model will give you the lowest bias:

- a) $\hat{f} = 3$
- b) $\hat{f} = w_0$
- c) $\hat{f} = w_0 + w_1x$
- d) $\hat{f} = w_0 + w_1x + w_2x^2$
- e) $\hat{f} = w_0 + w_1x + w_2x^2 + w_3x^3$

Which model will give you the lowest variance:

- a) $\hat{f} = 3$
- b) $\hat{f} = w_0$
- c) $\hat{f} = w_0 + w_1x$
- d) $\hat{f} = w_0 + w_1x + w_2x^2$
- e) $\hat{f} = w_0 + w_1x + w_2x^2 + w_3x^3$

3 Ridge Regression

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

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

Recall from the lecture that Ridge regression minimizes:

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

- a) By differentiating the loss function, show that Ridge regression has the following closed-form solution:

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

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

- b) What happens to the weights \mathbf{w}_{ridge}^* in the limit as $\lambda \rightarrow \infty$?
c) What happens to the weights \mathbf{w}_{ridge}^* in the limit as $\lambda \rightarrow 0$?

4 Rescaling

Suppose we have a dataset D with 1000 samples and 100 features $\{x_1, x_2, \dots, x_{100}\}$. 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 $w_1^{*'} = \frac{1}{10}w_1^*$
- 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?