

Large Scale Machine Learning and Deep Learning

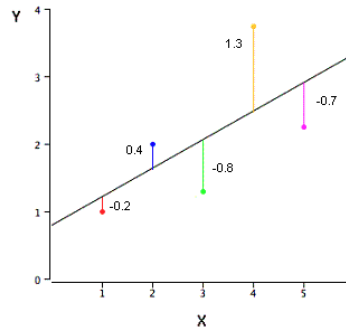
Review Questions 1

1. Which of the following is/are true about *Normal Equation*?

- (a) We don't have to choose the learning rate.
- (b) It becomes slow when number of features is very large.
- (c) No need to iterate.

Answer: a, b, c

2. The following graph represents a regression line predicting y from x . The values on the graph shows the residuals for each predictions value, i.e., $\hat{y} - y$. Calculate the squared error of the prediction.



Answer: $squared_error = (-0.2)^2 + (0.4)^2 + (-0.8)^2 + (1.3)^2 + (-0.7)^2 = 3.02$

3. How does number of observations influence overfitting? Choose the correct answer(s).

- (a) In case of fewer observations, it is easy to overfit the data.
- (b) In case of fewer observations, it is hard to overfit the data.
- (c) In case of more observations, it is easy to overfit the data.
- (d) In case of more observations, it is hard to overfit the data.

Answer: a and d. In particular, if we have very few observations and its small, then our models can rapidly overfits data. Because we have only a few points and as we're increasing in our model complexity like the order of the polynomial, it becomes very easy to hit all of our observations. On the other hand, if we have lots of observations, even with really complex models, it is difficult to overfit, because we have dense observations across our input.

4. How many coefficients do you need to estimate in a simple linear regression model (one independent variable)?

Answer: in simple linear regression ($\hat{y} = w_0 + w_1x_1$), there is one independent variable, so it has two coefficients, i.e., w_0 and w_1 .

5. Suppose you have fitted a complex regression model on a dataset. Now, you are using Ridge regression with tuning parameter lambda (λ) to reduce its complexity. Choose the option(s) below which describes relationship of bias and variance with lambda.

- (a) In case of very large λ ; bias is low, variance is low.
- (b) In case of very large λ ; bias is low, variance is high.
- (c) In case of very large λ ; bias is high, variance is low.
- (d) In case of very large λ ; bias is high, variance is high.

Answer: c. The large λ decrease the complexity of the model, and therefore the bias is high and variance is low.