

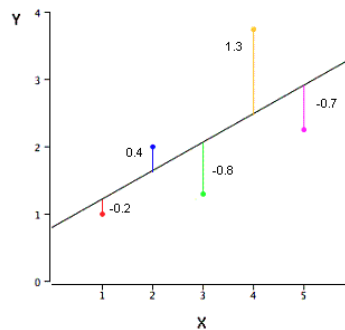
Scalable Machine Learning and Deep Learning - Review Questions 1

Deadline: Nov. 13, 2021

1. **0.5 point.** 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.
-

2. **0.5 point.** 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.



3. **0.5 point.** 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.
-

4. **0.5 point.** How many coefficients do you need to estimate in a simple linear regression model (One independent variable)?

5. **0.5 point.** What is cross validation and how does it work?

6. **1 point.** Mathematically show that the softmax function with two classes ($k = 2$) is equivalent to the sigmoid function?

7. **0.5 point.** As you know, in binomial logistic regression the **cost** between the true value y and the predicted value \hat{y} is measured as below:

$$\text{cost}(\hat{y}, y) = \begin{cases} -\log(\hat{y}) & \text{if } y = 1 \\ -\log(1 - \hat{y}) & \text{if } y = 0 \end{cases}$$

Explain why $-\log$ is a proper function to compute the cost in logistic regression?

8. **0.5 point.** How are logistic regression cost, cross-entropy, and negative log-likelihood related?
-

9. **0.5 point.** Explain how a ROC curve works?