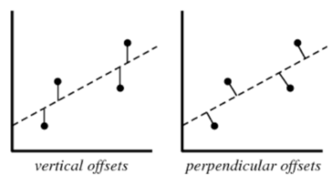
# Discussion Session Problems 1

01/14/2026

1. Which of the following offsets, do we use in linear regression's least square line fit? Assume the horizontal axis is the independent variable and vertical axis is dependent variable.




   A) Vertical offset
   B) Perpendicular offset
   C) Both, depending on the situation
   D) None of above

2. Which of the following if any is a valid cost function in a regression setting and why?

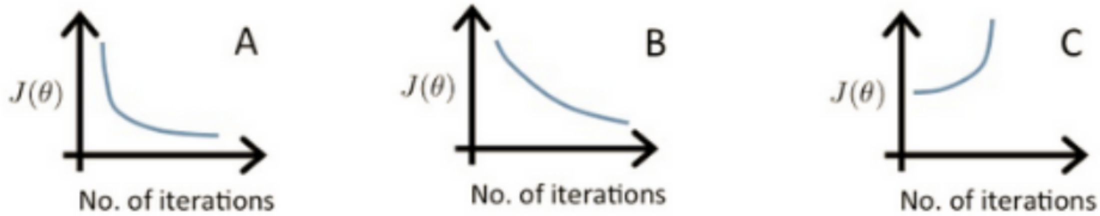
   a. $J(w) = \frac{1}{2m} \sum_{i=1}^{m} (f_w(x^{(i)}) - y^{(i)})^2$
   b. $J(w) = \frac{1}{2m} \sum_{i=1}^{m} (f_w(x^{(i)}) - y^{(i)})$
   c. $J(w) = \frac{1}{2m} \sum_{i=1}^{m} |f_w(x^{(i)}) - y^{(i)}|$

3. You are building a model to set real estate prices. If the predicted price is too high no customer will buy the house, but the monetary loss is low because the price can easily be decremented. Of course it should not be too high as then the house may not be bought for a long time. On the other hand if the predicted price is too low, the house will be bought quickly without having a chance to adjust the price. In other words the learning algorithm should predict slightly higher prices which can be decremented if necessary rather than underestimating the 'good' price which will result in an immediate monetary loss. How would you design an error metric incorporating this cost asymmetry? Write your new cost function and draw a sketch of the graph where you plot the cost versus $w$.

4. An outlier in a regression model:

   a) Always improves model accuracy
   b) Will most likely change or skew the regression line's slope
   c) Has no impact on the model
   d) Automatically indicates a data collection error

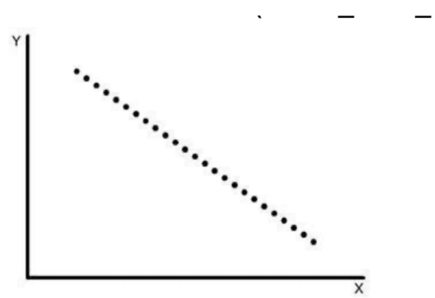
5. a. What can you say about the relationship between the cost function and the number of iterations in the graphs below?




b. Suppose $l_1$, $l_2$ and $l_3$ are the three learning rates for A, B, C respectively. Which of the following is true about $l_1$, $l_2$ and $l_3$?

A) $l_2 < l_1 < l_3$
B) $l_1 > l_2 > l_3$
C) $l_1 = l_2 = l_3$
D) None of these

6. Consider the following data where one input(X) and one output(Y) is given. What would be the cost for this data if you run a Linear Regression model of the form ($Y = w_1 \cdot x_1 + b$)?




A) Less than 0
B) Greater than zero
C) Equal to 0
D) None of these

| Hyperparameter setting | Learning Rate | Number of iterations | Training Error | Validation Error |
|---|---|---|---|---|
| 1 | 0.1 | 1000 | 100 | 110 |
| 2 | 0.2 | 600 | 105 | 105 |
| 3 | 0.3 | 400 | 110 | 110 |
| 4 | 0.4 | 300 | 120 | 130 |
| 5 | 0.4 | 250 | 130 | 150 |

7. Which of the following hyperparameter settings is seemingly the best?

A) 1
B) 2
C) 3
D) 4

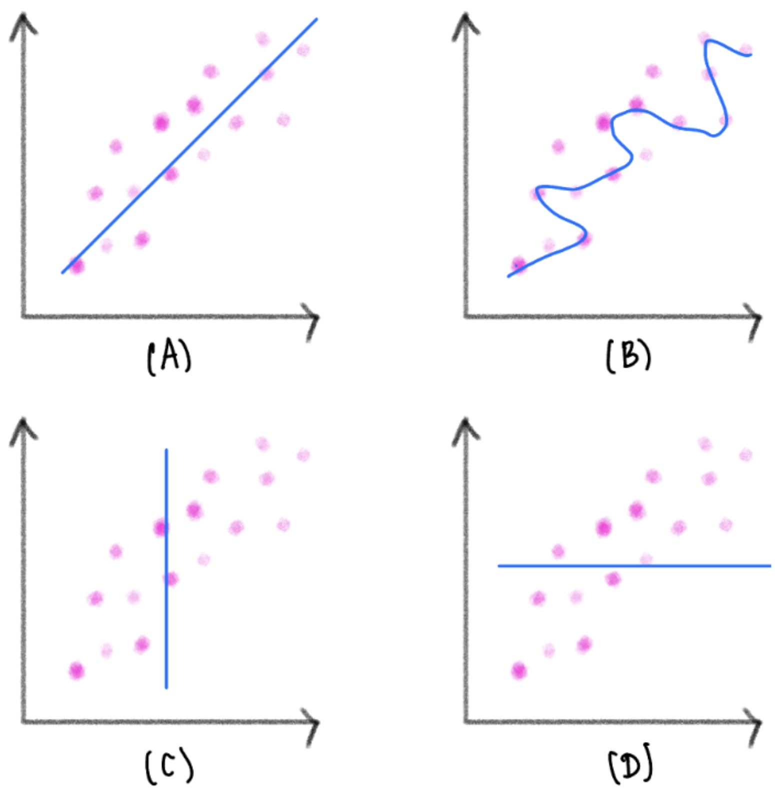
8. Normal equations are very slow when we have a big X. However technically they should work all the time. **True/False**

9. Why is feature scaling important in linear regression?

   a) It always improves model accuracy
   b) It prevents variables with larger magnitudes from dominating
   c) It guarantees perfect predictions
   d) It reduces computational time

10. Can we model non-linear relationships with a linear regression?

11. Which of the following is/are a theoretically possible linear regression fit? State how or why.




12. You want to extend your linear regression approach to capture a non-linear relationship by creating polynomial features (e.g., $x^2$, $x^3$, etc.).

    (a) How can this approach still be considered "linear" regression?
    (b) What potential problem might arise when adding too many polynomial terms, and how can you mitigate it?

13. Why do we split our dataset into training and test sets when building a regression model?