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 the vertical axis is dependent variable.

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

Solution: (A)
We always consider residuals as vertical offsets. We calculate the direct differences between actual value and the Y labels. Perpendicular offset are useful in case of PCA.

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

\[ J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 \]

\[ J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) \]

\[ J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} |h_\theta(x^{(i)}) - y^{(i)}| \]

A yes
B no - (how would you minimize it?)
C yes

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 theta.

Solution: penalize more undershooting than overshooting. Check out this link for a more detailed explanation.

Different methods:
- Square on the left/absolute value on the right
- Weigh more the left than the right like \( \frac{2}{3} / \frac{1}{3} \)
- Multiply the cost by some constant squared and vary that constant accordingly.

4) Which of the following statements is true about outliers in Linear regression?

A) Linear regression is sensitive to outliers
B) Linear regression is not sensitive to outliers
C) Can’t say
D) None of these
Solution: (A)
The slope of the regression line will change due to outliers in most of the cases. So Linear Regression is sensitive to outliers.

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

Figure A shows a just right learning curve. As you lower your learning rate you will get figure B. If you increase it you can end up overshooting and thus you will get figure C.

b. Suppose l1, l2 and l3 are the three learning rates for A, B, C respectively. Which of the following is true about l1, l2 and l3?

A) l2 < l1 < l3
B) l1 > l2 > l3
C) l1 = l2 = l3
D) None of these

A) In case of high learning rate, step will be high, the objective function will decrease quickly initially, but it will not find the global minima and objective function starts increasing after a few iterations.

In case of low learning rate, the step will be small. So the objective function will decrease slowly.

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 = theta_0 + theta_1 * x_1)?

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

Solution: (C)
We can perfectly fit the line on the following data so mean error will be zero.
Question Context 27-28:
Suppose you have been given the following scenario for training and validation error for Linear Regression.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Learning Rate</th>
<th>Number of iterations</th>
<th>Training Error</th>
<th>Validation Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
<td>1000</td>
<td>100</td>
<td>110</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
<td>600</td>
<td>105</td>
<td>105</td>
</tr>
<tr>
<td>3</td>
<td>0.3</td>
<td>400</td>
<td>110</td>
<td>110</td>
</tr>
<tr>
<td>4</td>
<td>0.4</td>
<td>300</td>
<td>120</td>
<td>130</td>
</tr>
<tr>
<td>5</td>
<td>0.4</td>
<td>250</td>
<td>130</td>
<td>150</td>
</tr>
</tbody>
</table>

7) From which of the following scenario would you choose the right hyper parameter?
   A) 1   B) 2   C) 3   D) 4

Solution: (B)
Option B would be the better option because it leads to less training as well as validation error.

8) Normal equations are very slow when we have a big X. However technically they should work all the time. **True (when X is not invertible you can use pinv)**

9) What happens in gradient descent when $\alpha < 0$?
   Gradient ascent

10) Can we model non-linear relationships with a linear regression?
    Yes → transform features

11) Classification/regression/classification