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.

![Image of vertical and perpendicular offsets]

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?

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

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

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

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

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

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 * x_1 + b)?

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

<table>
<thead>
<tr>
<th>Hyperparameter setting</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) 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) What happens in gradient descent when α < 0?

10) Can we model non-linear relationships with a linear regression?
11) Explain whether each of the following situations is a classification or regression problem:
- A company wants to launch a new product and wants to know whether it will turn out to be a success or failure. We have information on the last 100 products this company launched, including if it was a success/failure, price, weight, color, and several other variables.
- We have information on several Bay Area Tech Companies, including size, industry, revenue, average employee salary, and more. We want to know which features influence the average employee salary.
- You are given data of 100 individuals and their sequenced DNA and want to know whether these individuals will exhibit a particular disease based off their genomic mutations. We have information on 10,000 individual genomes and whether or not they exhibit the particular disease.