Module 4 Quiz

1. Which of the following are true about linear regression (select all that apply)?
   - Linear regression forms the basis of more complex machine learning models
   - Linear regression is a great first model to apply to a problem to get a baseline of expected performance
   - Linear regression models commonly overfit
   - Linear regression models are easy to interpret, relative to other types of models

2. How do we determine the optimal coefficient values in linear regression?
   1. We find the coefficient values which result in the minimum value of our cost function, the Cross Entropy
   2. We find the coefficient values which result in the maximum value of our cost function, the Sum of Squared Error (SSE)
   3. We find the coefficient values which result in zero error
   4. We find the coefficient values which result in the minimum value of our cost function, the Sum of Squared Error (SSE)

3. If we have a situation where we are seeking to model a target variable that has a nonlinear relationship with one of our input features, which of the following is true?
   1. We can use linear regression and improve the fit by transforming the input feature using a nonlinear function into a new variable which we then add into the linear regression model as a feature
   2. We cannot use linear regression
   3. We can only use neural networks in situations where we have nonlinear relationships between input features and the output target
   4. We can use linear regression but would get a better fit using a logistic regression model

4. Why do we often use regularization in modeling?
   1. It helps to reduce underfitting by encouraging the development of more complex models
   2. It helps to reduce overfitting by penalizing complexity of the model
   3. It usually improves the performance of the model on the training dataset
   4. It reduces the computational complexity of training our model

5. What is the key difference between LASSO and Ridge regression?
   1. LASSO regression is only used with logistic regression for classification while Ridge regression is only used with linear regression for regression tasks
   2. LASSO regression forces coefficients of irrelevant features to 0 while Ridge regression forces coefficients to be small but not to 0. Therefore LASSO can be considered a feature selection technique but not Ridge
   3. Ridge regression forces coefficients of irrelevant features to 0 while LASSO regression forces coefficients to be small but not to 0. Therefore Ridge can be considered a feature selection technique but not LASSO

4. LASSO regression adjusts the bias term but not the coefficients of the model, which Ridge adjusts both

6. Which of the following are correct about linear regression and logistic regression (select all that apply)?

- Linear regression attempts to predict the output value, while logistic regression predicts the probability that the output is the positive class
- Both linear regression and logistic regression are used as algorithms for modelling regression tasks
- Both linear regression and logistic regression are linear models
- Linear regression is used for modelling regression tasks while logistic regression is used for classification tasks

7. Why is the logistic (or sigmoid) function used in logistic regression?

1. We use the logistic function to convert the numerical output into a probability of the postive class, in the range of 0 to 1
2. The logistic function is used to create a nonlinear combination of the input variables
3. The logistic function is used as regularization of the logistic regression model
4. We use the logistic function to convert the numerical output into a probability of the negative class, in the range of 0 to 1

8. Which of the following are correct regarding the use of gradient descent in logistic regression (select all that apply)?

- We only use gradient descent to find the coefficients/weights in neural networks, but not for logistic regression
- During each iteration of gradient descent, we adjust the coefficient/weight values in the direction opposite the gradient of the cost function with respect to the weights
- Gradient descent is an iterative solving method that helps us find the coefficients/weights which minimize the cost function
- Before we begin to perform gradient descent, we initialize the coefficients/weights of our model equation to small random values

9. We are building an app that uses pictures which our users upload from their phones to identify species of birds in the photos. We want to be able to identify 20 different types of birds with our model. Which of the below algorithms might be the best choice to use for our model?

1. Softmax regression
2. Linear regression
3. None of the above
4. Standard logistic regression (non-softmax)

10. Which of the following are true about softmax regression?
    - It is used for classification tasks where we have more than 2 possible output classes we are trying to predict
    - It is used primarily for binary classification tasks
    - The sum of the outputs from a softmax regression model across all classes is 1
    - We take the class with the lowest corresponding output value from the softmax regression model as the predicted class