UNIT 4 ASSIGNMENT

Introduction to Linear Models

## Instructions

The questions below will prepare you for future interviews as they relate to concepts discussed throughout the unit. You’ve practiced these concepts in the coding activities, exercises and coding portion of the assignment. Now, let’s formulate your programming into well-thought responses.

Except as indicated, use this document to record all your assignment work and responses to any questions. At a minimum, you will need to turn in a digital copy of this document to your facilitator as part of your assignment completion. You may also have additional supporting documents that you will need to submit. Your facilitator will provide feedback to help you work through your findings.

**Note:** Though your work will only be seen by those grading the course and will not be used or shared outside the course, you should take care to obscure any information you feel might be of a sensitive or confidential nature.

*Begin your assignment by completing the questions below. Directions to submit your work can be found on the assignment page. Information about the grading rubric is available on any of the course assignment pages online. Do not hesitate to contact your facilitator if you have any questions about the assignment.*

Unit 4 Written Portion

# Logistic Regression

Answer the questions below about linear models.

## Questions:

1. What is a linear model? What are the advantages and disadvantages of linear models?

|  |
| --- |
| A linear model is a supervised learning models that use the linear combination of features and weights to compute an unlabeled example. The advantages of a linear model is that they can improve themselves during training by using an optimization technique known as loss function to minimize training loss. They also use regular ideation to help reduce model complexity. The disadvantage of linear model is that it’s not ideal for a a model that’ll have a curve. It assumes that the data relationship is well described by a straight line. |

1. What type of supervised learning problem is logistic regression best suited for? Give an example of a problem you would use a logistic regression model for. Explain what you are trying to predict.

|  |
| --- |
| The type of supervised learning problem that a logistic regression is best suited for is classification problems. Specifically binary classification problems. An example in which I would use a logistic regression model for is predicting the probability of a class label such as probability that an email is spam or non-spam. |

1. Describe the training phase of a logistic regression model: explain the intuition behind using gradient descent algorithm and the use of loss functions.

|  |
| --- |
| The training phase of a logistic regression model is done iteratively with the help of the loss function and the optimization algorithm—the gradient descent—which uses the loss function to evaluate a model’s loss and then adjusts the model parameters accordingly to reduce loss. The loss function is used to evaluate the performance of the model against training data at any point in time. The most common loss function we use is the log loss aka. Cross entropy. In each iteration we evaluate the loss of our model against a set of training examples, this would allow us to get a performance of our logistic regression model with the weights and the intercept at that particular time. Our gradient descent updates the weights and intercepts on poorly performing models by starting with an arbitrary value for model parameters, then take the partial derivative of the loss function, then update the model parameters. We use a formula to compute the next value of weights and intercepts. The goal is to update the value of W to reduce loss until the slope is close to 0. |

1. Explain the purpose of using regularization when training a logistic regression model.

|  |
| --- |
| The purpose of regularization is to help minimize loss and to avoid overfitting by minimizing the model’s complexity. It’ll penalize complex models to prevent overfitting. |

1. Explain which linear model and accompanying loss function you would use for a classification problem and for a regression problem.

|  |
| --- |
| Logistic regression is used for classification problem and the loss function we would use is the log loss to measure the performance of a regression model. For linear regression it’s used for regression problems and we can use the mean square error which is used to measure the performance of a regression model. |

*To submit this assignment, please refer to the instructions in the course*.