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 week. 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 project 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 project 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.

*Complete each project part as you progress through the course. Wait to submit the project until all parts are complete. Begin your course project by completing Part One below. A submit button can be found on the final Course Project assignment page. Information about the grading rubric is available on any of the course project assignment pages online. Do not hesitate to contact your facilitator if you have any questions about the project.*

Week 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 any model that assumes there is a linear relationship between features (X) and the label (Y). Linear models are fairly simple so they require less data, are easy to interpret and are less prone to over-fitting. At the same time, they can not capture complex patterns in the data so may not perform as well as other methods. |
| --- |

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 (what are you trying to predict).

| Logistic Regression is designed for classification problems. An example classification problem is predicting whether or not a customer will churn from a subscription service. |
| --- |

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

| Training a Logistic Regression involves computing the log-loss and then adjusting the model weights to get to a point of lower loss. This is performed iteratively until the log-loss reaches a minimum point. Gradient Descent is an algorithm that defines how the model weights get adjusted at each iteration. If we thought of the loss-function as having a bowl shape, and a marble dropped in the bowl as our current model weight, gradient descent shows us the direction that gravity would pull the marble. Eventually the marble finds the bottom point of the bowl. |
| --- |

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

| Regularization is a way to constrain the flexibility of the logistic regression algorithm. The complexity of logistic regression comes from the features, and regularization limits either how many features have non-zero weights or the sensitivity of the prediction to the feature. This is ultimately used to control over-fitting. |
| --- |

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

| We would use Logistic Regression with Log-Loss for classification and Linear Regression with Mean Squared Error for regression problems. |
| --- |

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