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?

| Linear models are a class of supervised learning models that are represented by an equation and use a linear combination of features and weights to compute the label of an unlabeled example. The advantages: they’re simple to implement, fast to train, lower in complexity, good with smaller data. Disadvantages: too simple so it doesn’t adapt to nuances as well, assumes linearity. |
| --- |

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.

| Logistic regression is best suited for binary classification problems. It is used to estimate the probability that a new, unlabeled example belongs to a given class. One example commonly used example is whether an email is spam or not. As it trains, it adjust the weights for the best chance of a prediction. It’s pretty good for interpretability because you’re manually changing the coefficients. |
| --- |

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

| When training a logistic regression model, you use gradient descent and loss functions. The loss functions evaluate a model on the training data and tell us how bad the performance is. A loss of 0 means the model makes perfect predictions; higher = worse. The Gradient descent is an iterative optimization algorithm. It minimizes the loss function and finds the optimal weights. The weights are updated and multiplied by a learning rate to approach the minimum of the loss function. |
| --- |

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

| Regularization prevents overfitting. In the Log regression model, the C is what controls the complexity. If you have a higher C, it’s less regulation and creates a higher model complexity. So, you should try different C to avoid overfitting. It also helps by putting a boundary on the model weights → doesn’t overfit. |
| --- |

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

| For classification problems, logistic regression can be used for binary classification tasks because it deals directly with the probability of an input belonging to a particular class  The loss functions should be the log loss, which measures the performance and similarity between the predictions and true labels and then penalizes or rewards the system based on outcomes.  For regression, a linear model would be best because it models relationships between the inputs and a continuous variable. The model assumes a linear relationship and finds the best line to minimize the difference between predictions. Mean Squared Error is used for loss function most often because it takes the difference between the label and the prediction and squares it. |
| --- |

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