ECE M148 Homework 4

Damien Ha, UID 905539967

Question 1

Part A

False Positive: $\frac{10}{25+10} = \frac{10}{35} = \frac{2}{7}$

False Negative: $\frac{9}{36+9} = \frac{9}{45} = \frac{1}{5}$

Part B

If we increase t, then less samples will be classified as positive. Both true positives and false positives would then decrease, while both true negatives and false negatives would increase.

Problem 2

A: fire occurs in kitchen B: smoke detector sets off alarm

P(A) = 0.01 P(B|A) = 0.99 P(B|A') = 0.10

Part A

$$P(B) = P(B \mid A) * P(A) + P(B \mid A') * P(A')$$

$$= 0.99 * 0.01 + 0.10 * (1 - 0.01)$$

$$= 0.0099 + 0.099$$

$$= 0.1089$$

Part B

$$P(A \mid B) = \frac{P(B|A) * P(A)}{P(B)}$$

$$= \frac{0.99 * 0.01}{0.1089}$$

$$= \frac{0.0099}{0.1089}$$

$$\approx 0.0907$$

Part C

Admittedly this smoke detector has a high probability of sounding an alarn when there is a fire at 99%, but it also has a somewhat high rate of giving alarms when there is not a fire as we see in part b. Given this, it might be worth replacing this alarm that does not have as high a rate of false positives if we assume false fire alarms are considered a major inconvenience.

Problem 3

$\frac{dL(\beta)}{d\beta_j} = \sum_{i=1}^{n}(\frac{1}{1+e^{-\beta^T x_i}} - y_i)x_{ij}$

$L(\beta) = -\sum_{i=1}^{n}(y_i log(1+e^{-\beta^T x_i}) + (1-y_i)log(\frac{1}{1-e^{-\beta^T x_i}}))$

$\frac{dL(\beta)}{d\beta_j} = -\sum_{i=1}^{n}(y_i \frac{d}{d\beta_j} log(1+e^{-\beta^T x_i}) + (1-y_i)\frac{d}{d\beta_j} log(\frac{1}{1+e^{-\beta^T x_i}}))$

and we get $= \sum_{i=1}^{n}((\frac{1}{1+e^{\beta^T x_i}})x_{ij})$

Because $\frac{d}{d\beta_j} log(1+e^{-\beta^T x_i}) = -x_{ij}(\frac{1}{1+e^{\beta^T x_i}})$

and $\frac{d}{d\beta_j} log(1-e^{-\beta^T x_i}) = x_{ij}(\frac{1}{1+e^{\beta^T x_i}})$

Problem 4

Part A

In one-vs-all, a binary classification model is trained for each class seperately, testing if it is in or not in the class

Advantages: It's simple and easy to implement, and can handle class imbalance

Disadvantages: It's complex when there is a large number of cases and doesn't handle cases with overlap between classes well

Part B

In all-vs-all, a binary classificantion model is trained on every pair of classes so it can distinguish between every possible pair.

Advantages: It handles cases with overlap between classes well, better than one-vs-all, and can be better for large numbers of cases

Disadvantages: It is computationally expensive due to training classifiers for every single pair

Problem 5

Part A

True. Positive predicted values is true positive predictions over total positive predictions, which is the probability a sample is actually positive given that the model classified it as positive

Part B

False, for multinomial logistic regression with a dataset of 1 feature and K possible class labels, the number of learnable parameters $B_{j,i}$ is K-1. In this case this would be 4 - 1 = 3

Part C

True, when modeled as a quadratic it is not linear so it allows for a non-linear decision boundary

Part D

False. Test accuracy alone is not enough, in this case Type I error (rate of false positives) is not considered

Part E

True, the confusion matrix will have dominant values on the diagonal when the model is correctly predicting instances of each class