COMPSCI 570 **Instructor:** Ronald Parr

Homework 3 **Due:** 11/5/24

Note: You may **not** search the internet, use ChatGPT/ChatGPT-like systems, or use any other (generative)AI/foundation/large-language models to help answer these questions. You may use the internet to look up basic definitions and methods relating to probability, expectation, and Bayesian probability.

You may have high level discussions with your classmates about understanding the underlying concepts and about how to approach these problems in general, but may not share anything resembling a complete solution. What you write up and turn in should be your own.

1 Probabilities in Logistic Regression (25 points)

Derive a relationship between the distance a point is from the decision boundary in logistic regression and the probability of the label assigned by logistic regression.

2 Naive Bayes (25 points)

It's possible to show that any Naive Bayes classifier defined over m binary random variables can be converted to a linear threshold defined over m binary features and a constant. Your answer should show mathematically how the CPTs from the Naive Bayes classifier are converted to weights. **Hint:** In naive Bays, we are concerned about comparing the probabilities of being positive or negative, $P(y|x)$ and $P(\overline{y}|x)$ for any data point. Suppose we want to show $P(y|x) > P(\overline{y}|x)$, meaning that the data point is more likely to be positive, how should you convert that into the form of a binary classifier, $W^T x + b > 0$?

3 Linear Regression I (25 points)

Provide math derivations for the following questions.

1. Assume w is the least squares regression solution for feature set Φ, with first feature $\phi_1 = 1$, i.e., a constant. Now suppose we create a new feature set Φ' such that all features are the same except for a single ϕ_i ($i > 1$) such that $\phi'_i = \phi_i + c$, where c is an arbitrary constant. How does that change the optimal solution, i.e., if w' is the solution Φ', what is the relationship between w' and w? (Be sure to justify why your proposed new optimal solution is indeed optimal.) (15 points)

2. Suppose w is the least squares regression solution for feature set Φ and targets y, with mean squared error $e = \|\Phi w - y\|_2^2/n$. If we scale both the targets and the features by k, i.e., $\Phi' = k\Phi$, and $y' = ky$, what is the relationship between $e' = \|\Phi' w' - y'\|_2^2/n$, and e?

4 Linear Regression II (25 points)

Assume your features can be divided into two sets, Φ_1 and Φ_2, and that all features in Φ_1 are orthogonal to features in Φ_2. Prove that the ordinary least squares regression for target X with weights w has the following form:

$$w = \begin{pmatrix} (\Phi_1^T \Phi_1)^{-1} & 0 \\ 0 & (\Phi_2^T \Phi_2)^{-1} \end{pmatrix} [\Phi_1 \Phi_2]^T X$$

If you are stumped on this, try Googling: block diagonal matrix inversion. Note that we are using the notation: $[\Phi_1 \Phi_2]$ to indicate the result of concatenating Φ_1 and Φ_2 to form one big matrix.

Why this is an interesting observation: If you have two sets of orthogonal features, then it's like having to independent regression problems that can be solved separately.