1 Linear Separability (10 points)

Suppose function \( f \) is \textit{not} linearly separable in the feature space \( \phi_1 \ldots \phi_k \), but you are hoping that it will be separable when you add a new feature \( \phi_{k+1} \). It turns out, however, that your choice of \( \phi_{k+1} \) can be expressed as a linear combination of \( \phi_1 \ldots \phi_k \). Prove rigorously that \( f \) must not be separable in the new feature space, \( \phi_1 \ldots \phi_{k+1} \).
2 Probabilities in Logistic Regression (10 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.
3 Naive Bayes (10 points)

In class (and slides), we made the comment that naive Bayes was also a linear method, but this wasn’t fully fleshed out. Suppose you have a naive Bayes classifier with $n$ binary features $\phi_1 \ldots \phi_n$ and a binary class label $t$. Derive a linear classifier from the probabilities in the Bayesian network where the classifier is defined over a set of binary inputs. Note that this isn’t entirely trivial because the Bayesian network will have four parameters per feature, but your classifier will only have one per feature. It’s OK to assume that probabilities aren’t 0.
4 Linear Discriminant Analysis (10 points)

A peculiar property of linear discriminant analysis is that it’s possible for the center (mean) of both Gaussians to be on the same side of the decision boundary. Provide a one-dimensional example of how this is possible. Start by simplifying the form given in class for the one-dimensional case (5 points), then provide the necessary parameters and prove that your example has both means on the same side of the decision boundary (5 points). (Note that n in the slides is the dimension of the Gaussian, which is the same as the number of features.)