Assignment 4

Written Problems

This part is due back in class on Thursday November 12.

1. Read problem 4.9 page 221 and its solution (on text web page). Then solve problem 4.10 using the same representation for the multiclass problem.

2. Calculate one iteration of the Newton-Raphson method for minimizing the function \( f(x) = x_1^3 + 5x_1x_2^2 - 7x_1^2x_2 \) where \( x = (x_1, x_2)^T \). Use \( x = (1, 1)^T \) as the initial value.

3. Complete the details from class (and text) and give the iterative update formula for calculating \( w_{MAP} \) for the Bayesian logistic regression. You may quote any results and formulas you need from class or textbook (cf. page 218).

4. Develop a Gaussian approximation to the beta distribution \( \text{beta}(\mu|a, b) \) using the Laplace approximation. First develop the formulas in general and then apply them to the case \( a = b = 3 \).

5. Solve problem 6.8 page 320.


Experiments with Classification Algorithms

This part may be submitted separately, by Tuesday November 17th in class.

On the course page you will find 3 datasets for experimentation with classification algorithms. The first dataset (marked as A) has uniformly distributed data with an arbitrary weight vector separating positive from negative examples. Therefore data does not conform to the Gaussian generative model but the data is linearly separable. For the second dataset (marked as B) data is generated using the Gaussian generative model where the two classes have different covariance matrices and there is some overlap between the distributions. The third dataset, ionosphere, is taken from the UCI repository. Each dataset is given in two files with the data in one and the labels in the other file. Unlike previous assignments I did not split the data into a training set and test set and you will use cross validation to report results.

Your task in this assignment is to implement 4 different learning algorithms and evaluate their performance on these datasets. In particular, you should implement the (1) 2 class generative model with shared covariance matrix, (2) 2 class generative model with separate a covariance matrix for each class, (3) Fisher’s linear discriminant with threshold given by \( w^T \cdot m \) where \( m \) is
the mean of the examples, (4) Bayesian logistic regression where you can pick the regularization parameter arbitrarily. Note that unlike the other algorithms, logistic regression relies on a free parameter \( w_0 \) to capture an appropriate separating hyperplane. Therefore, you will need to add a feature fixed at one to the data for this algorithm.

For each dataset, perform 10-fold cross validation and report the obtained accuracies and standard deviations. Repeat this for different training set sizes within each fold to obtain learning curves (with error bars) for the algorithms.

Please submit printouts of your code, and a short report on the experiments, their results, and your conclusions from them.