Statistical Learning-Classification
STAT 841 / 441, CM 763

Assignment 2

Department of Statistics and Actuarial Science
University of Waterloo

Due: Thursday October 26 at 1:00 PM

Policy on Lateness: Slightly late assignments (up to 24 hs after due date) are accepted with 10% penalty. No assignment are accepted after 24 hs after the due date.

Note: Matlab is not mandatory. You can use any programming language.

1. **Face detection**:
   Download the faces.mat from the course webpage.
   The file *faces.mat* is composed of *train_faces, train_nonfaces, test_faces*, and *test_nonfaces*.
   Make a training and a test set as follows:

```
training_data=[train_faces' train_nonfaces'];
% (This will be a 361 by 4858 matrix.)

test_data=[test_faces' test_nonfaces'];
% (This will be a 361 by 944 matrix.)
```

   a) Write a program to fit a logistic regression model to the training data. Report the first 5 components of the optimum value of the logistic parameter β, as well as the training error and the test error.

   b) Write a program to fit a single hidden layer neural network with 4 hidden units and one output node via back-propagation. Report the training error and the test error.

   Note : Attach your code to your assignment as an appendix, and submit the code to the assignment drop box in Learn as well.

2. In a maximum likelihood problem, we can define an error function by taking the negative logarithm of the likelihood. Show that the error function for the logistic regression model is a convex function of β, and hence show that it has a unique minimum value.

3. Consider a neural network that consists of a single neuron with d inputs. The neuron has d weights, $w \in \mathbb{R}^d$. The output of the neuron for an input pattern $x \in \mathbb{R}^d$ is given by $\hat{y} = \Phi(x \cdot w)$, where Φ is an activation function.

   For any differentiable activation function Φ, there exists a *matching loss*, denoted by $err_\Phi(y, \hat{y})$, such that when using Φ and its matching loss $err_\Phi(y, \hat{y})$, the error function of a single neuron is convex and thus has only one minimum. The *matching loss* can be computed as:

$$err_\Phi(y, \hat{y}) = \int_{\Phi^{-1}(y)}^{\Phi^{-1}(\hat{y})} (\Phi(z) - y) dz$$

   a) Find the matching loss for the activation function $\Phi_1(z) = z$.

   b) Find the matching loss for the activation function $\Phi_2(z) = \frac{1}{1+e^{-z}}$.

   c) Suppose you want to use this simple network for a classification task. Which of these two loss functions (Φ_1 or Φ_2) is more appropriate? Briefly explain why.

Only for Grad Students

4. Consider a multiclass logistic regression model (multilogit model) applied to d-dimensional data with K classes. Let β be the $(d + 1)(K - 1)$-vector consisting of all the coefficients. Define a suitably enlarged version of the input vector x to accommodate this vectorized coefficient vector. Derive the Newton-Raphson algorithm for maximizing the log-likelihood, and describe how you would implement this algorithm.

5. Consider a classification model for two classes with prior class probabilities $\pi_k, k = 1, 2$. Suppose that the class-conditional densities are given by Gaussian distributions with a shared covariance matrix. Suppose we are given a training data set $\{(x_i, y_i)\}$ where $i = 1 \ldots n$, and $y \in \{0, 1\}$ are class labels. Assume that the data points are drawn independently from this model.

   a) Compute the maximum-likelihood estimation for the prior probabilities.

   b) Compute the maximum-likelihood estimation for the mean of the Gaussian distribution for each class.

   c) Compute the maximum-likelihood estimation for the shared covariance matrix.