Selected Topic from Computer Science
Assignment 1

To be submitted before 11 pm on 2/03/18

The goal of this assignment is to implement different linear models for binary classification. Models should be implemented in C++, C or Java only, without the aid of any machine learning package. Below are the models you are expected to implement.

1. Fisher's linear discriminant- as described in Bishop ch 4. Further to find $y0$ (threshold) use minimum entropy criterion. After finding w where $w \propto S_w^{-1}(m_2 - m_1)$. Compute $w^T x$ (x is the input vector) for all training instances, the points are transformed into 1-dimension, sort the points. Lets say we have 5 instances and their transformed values are r_1, r_2, r_3, r_4 and r_5 ($r_j = w^T x_i$) where $r_1 > r_2 .. > r_5$. Now to find threshold y_0.

   for i *in* $1 \rightarrow 4$ **do**
   $f_i = \frac{r_i + r_{i+1}}{2}$
   compute entropy with f_i as threshold
   end for
   y_0 *is equal to* f_i *that gives minimum entropy*

   refer Tom Mitchell, Machine Learning ch 3 for a detailed description of entropy.

2. Probabilistic Generative Model-as described in Bishop ch 4
   - assume the class-conditioned densities $p(x|C_k)$ are Gaussian with both classes sharing the same covariance matrix
   - assume the posterior probabilities is given by a sigmoid function

3. Logistic Regression (Probabilistic Discriminative Model)-as described in Bishop ch 4.
   - Use gradient descent algorithm.(without using any package)
   - Experiment with different learning rates.

Dataset and Report
Use the two datasets uploaded on CMS. Use train.txt(960 instances) to train the model and test.txt(412 instances) to test model performance. The dataset was taken from uci repository's 'banknote authentication Data Set'. Each data instance has five values.

1. variance of Wavelet Transformed image (continuous)
2. skewness of Wavelet Transformed image (continuous)
3. curtosis of Wavelet Transformed image (continuous)
4. entropy of image (continuous)
5. class (0 or 1).

Four continuous attributes and a class value. The report should contain confusion matrix, recall and precision over testing data for all three models. **Upload the report and the working code file/s(in .txt format) to CMS**.

For queries contact:
Jinu Sunil
email id: f20130423@hyderabad.bits-pilani.ac.in