UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
Department of Electrical and Computer Engineering

ECE 544NA Pattern Recognition

Homework 4

Fall 2013

Assigned: Tuesday, September 24, 2013 Due: Thursday, October 03, 2013

Reading: NNPR Chapter 3

Problem 4.1

The hinge loss is a shifted version of the perceptron criterion, and is given by:

$$L(y, f(x)) = max(0, 1 - yf(x))$$

As discussed in class, this is now an upper bound on the 0-1 loss. Find the optimal classification (Bayes) rule given the hinge loss (i.e. find the optimal $f^*(x) = \arg\min_{f(x)} E_{X,Y}[L(y, f(x))]$ when $L(y, f(x))$ is the hinge loss). How is it related to the Bayes optimal rule for the 0-1 loss? **Note:** assume binary classification: $y \epsilon \{1, -1\}$

Matlab Exercises

Problem 4.2

Features such as plasma glucose concentration, blood pressure, body mass index, and age are used to diagnose a patient for diabetes. Since the unit of measurement across the features varies drastically, and there is no reason to believe (*a priori*) that certain features are more relevant than others, we normalize across all features.

Download the file diabetes_normalized from the class website. The first column denotes the class label, followed by eight normalized features relevant to the classification of diabetes. Each row is a sample point. For example, the first line **-1: 1:-0.294118 2:0.487437 ... 8: -0.0333** indicates that the *class* is -1, the *first* feature is -0.294118, the *second* feature is 0.487437, and so on, until the *eighth* feature (-0.0333). Randomly partition the dataset into two groups: **training set** and **test set**, where the **training set** contains **80 %** of the samples and the **test set** contains the remaining **20 %**. For each of the following scenarios, train a classifier on the **training set** and test it on the **test set**. Report the **average classification error** on the **test set**.

Note: As always, you are expected to submit your own matlab code and function(s) for each of parts (a) - (d).

(a) k-nearest neighbors classifier with the euclidean distance (l_2) for $k = 1$.

(b) Least-squares linear classifier.

(c) Non-linear activation function with $h(.) = tanh(.)$. Comment on your selection of η and the stopping criterion.

(d) The perceptron criterion (hinge loss). Is the training set linearly separable?

(e) Comment on your results, not only comparing the classification error of the classifiers, but also their fundamental approach, assumptions, and the advantages/disadvantages thereof. For example, c) and d) are iterative algorithms that rely on a stopping criterion – how does this influence the **test set** performance?