ECE523: Engineering Applications of Machine Learning and Data Analytics
Due 02/10/2017 @ 11:59PM (D2L)

Name: ______________________________

Signature: ______________________________

Date: ______________________________

Instructions: There are seven problems. Partial credit is given for answers that are partially correct. No credit is given for answers that are wrong or illegible. All work must be supported and code must be submitted for credit.

Theory: ___________

Practice: ___________

Total: ___________

Part A: Theory (15pts)

(5pts) Linear Regression and Regularization

In class we derived and discussed linear regression in detail. Find the result of minimize the loss of sum of the squared errors; however, add in a penalty for an L_2 penalty on the weights. More formally,

$$\arg\min_{\mathbf{w}} \left\{ \sum_i (\mathbf{w}^\mathsf{T}\mathbf{x}_i - y_i)^2 + \lambda \|\mathbf{w}\|_2^2 \right\} \tag{1}$$

How does this change the solution to the original linear regression solution? What is the impact of adding in this penalty?

(5pts) Density Estimation

In k-nearest neighbors (KNN), the classification is achieved by majority vote in the vicinity of data. Suppose there are two classes of data each of $n/2$ points overlapped to some extent in a 2-dimensional space. Describe what happens to the training error (using all available data) when the neighbor size k varies from n to 1.

(5pts) Feature Selection & Preprocessing

A friend asks you for some help with a feature selection project. Your friend goes out and collects data, $\mathscr{D}$, for their project. Using $\mathscr{D}$, your friend tries many subsets $\mathscr{F} \subset \mathscr{X}$ by adapting $\mathscr{F}$ based on the error of a classifier. They return $\mathscr{F}$ that corresponds to the smallest classification error.

This is the procedure they carry out to validate the impact of the feature selection routine. This procedure is repeated

- Make a new data set $\mathscr{D}'$ with $\mathscr{F}$ features using the feature selection routine.
- Repeat 50 times
  - Split $\mathscr{D}'$ into randomized training & testing sets (80/20% splits)
  - Train a classifier and record its error
- Report the error averaged over 50 trials

Critique and respond to how your friend performed their analysis.

Part B: Practice (25pts)

You are free to use functions already implemented in Matlab, Python or R with the exception of problem 1. I recommend using Python's Scikit-learn (`http://scikit-learn.org/stable/`) as is implements most of the methods we will be discussing in this course... as well as problems in this homework!

(10pts) Logistic Regression on Synthetic and Real-World Data

Write your own implementation of logistic regression and implement your model on either real-world (see Github data sets), or synthetic data. If you simply use Scikit-learn's (R's, Matlab's, or another builtin procedure) implementation of the logistic regression classifier then you'll receive a maximum of 6 out of 15 points. A full 10/10 will be awarded to those that actually implement logistic regression using the optimization of cross-entropy using stochastic gradient descent.

(5pts) Dimensionality Reduction + 2 Bonus

Choose 10 data sets of your choice from the ECE523. Implement a comparison between either two classifiers of your choice, or a classifier with does and does not using a preprocessing step (i.e., feature selection, PCA, etc.), and report the accuracies of the two models in table for (e.g., a 10×2 table of classifier accuracies). For two bonus points use a hypothesis testing procedure from homework #1 to determine if there is statistical significance (i.e., do both approaches perform equally well).

(10pts) Density Estimation in Practice

The ECE523 Github page has code for generating data from a checkerboard data set. Generate checkerboard data from two classes and use any density estimate technique we discussed to classify new data using

$$\widehat{P}(Y|X) = \frac{\widehat{P}(X|Y)\widehat{P}(Y)}{\widehat{P}(X)}$$

where $\widehat{P}(Y|X)$ is your estimate of the posterior given you estimates of $\widehat{P}(X|Y)$ using a density estimator and $\widehat{P}(Y)$ using a maximum likelihood estimator.