Computational Data Analysis

ISYE 6740

Final Exam– Due Dec. 13, 2019

Total Score: 100

If you think a question is unclear or multiple answers are reasonable, please write a brief explanation of your answer, to be safe. Also, show your work if you want wrong answers to have a chance at some credit: it lets us see how much you understood.

(Please sign the honor code below.) I have neither given nor received any unauthorized aid on this exam. I understand that the work contained herein is wholly my own without the aim from a 3rd person. I understand that violation of these rules, including using an authorized aid or copying from another person, may result in my receiving a 0 on this exam .


Name:

GT ID:

GT Account:

|	
Question 1 [15 points]	
Question 2 [15 points]	
Question 3 [15 points]	
Question 4 [20 points]	
Question 5 [10 points]	
Question 6 [25 points]	

1   Clustering [15 pts]

1. (5 points) Explain what is the difference between K-means and spectral clustering?

2. (5 points) What is in common, and what is the main difference between spectral clustering and PCA?

3. (5 points) For the following data (two moons), give one method that will successfully separate the two moons? Explain your rationale.
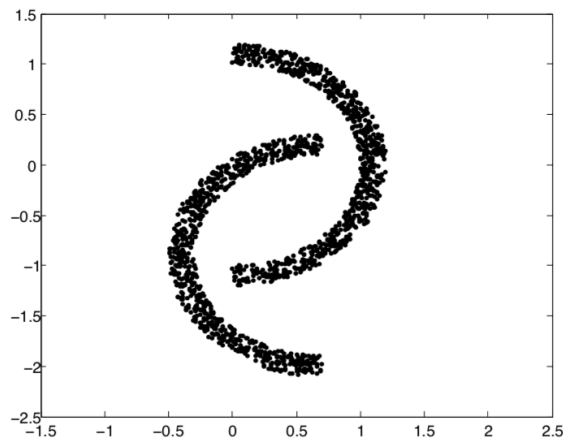
2 Classification [15 points]

1. (5 points) List all methods below which can be used for classification:

   (a) AdaBoost (b) Decision Trees (c) EM and Gaussian Mixture (d) Histogram (e) K-nearest neighbors (f) K-means (g) Kernel density estimation (h) Linear Regression (i) Logistic Regression (j) Naive Bayes.

2. (5 points) Which of the decision boundaries below correspond to (a) Random Forest, (b) Decision Tree, (c) SVM. Explain your reasons to fully justify your answers.
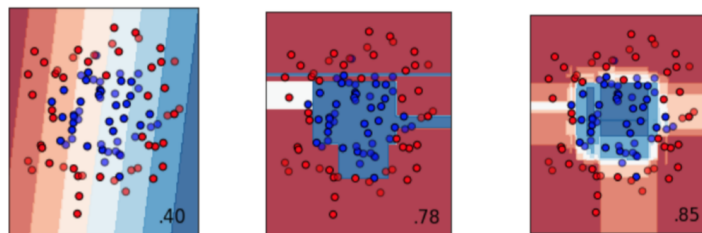
3. (5 points) Is the following statement true / false: "In the AdaBoost algorithm, the weights on all the misclassified points will go up by the same multiplicative factor." Explain your reason.

3   SVM [15 points]

Suppose we only have four training examples in two dimensions as shown in Fig. The positive samples at $x_1 = (0,0)$, $x_2 = (2,2)$ and negative samples at $x_3 = (h,1)$ and $x_4 = (0,3)$.
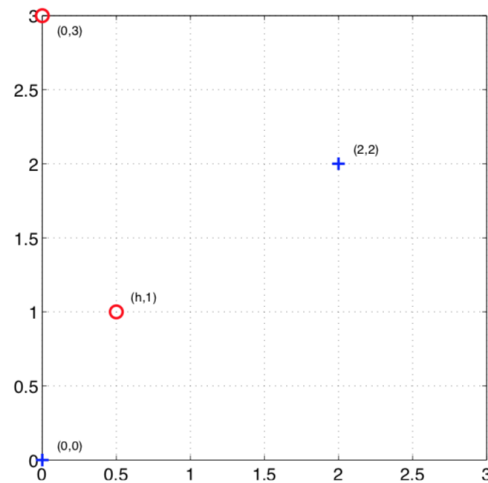

1. (5 points) For what h s.t. $h > 0$ to be so that the training points are still linearly separable?

2. (5 points) Does the orientation of the maximum margin decision boundary change as a function of h when the points are separable?

3. (5 points) Explain why only the data points on the "margin" will contribute to the decision boundary?

4   Variable section [20 points]

Suppose we have data $\{x_i, y_i\}$, $i = 1, \ldots, m$, where $x_i \in \mathbb{R}^p$ corresponds to p features.

1. (5 points) Write down the optimization problem we solve with Ridge Regression and Lasso. Make sure you explain your notations: which are the decision variables, and which are data.

2. (5 points) Which of the solution paths below corresponds to Ridge regression and which corresponds to Lasso?
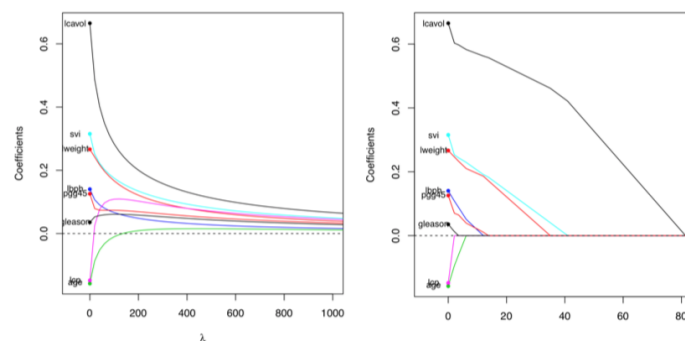
3. (5 points) Explain what's the difference between Lasso and Ridge regression. We need Lasso for what setting?

4. (5 points) Explain how to tune the regularization parameters for Lasso and Ridge regression.

5   Neural networks [10 points].

1. (5 points) Consider a neural networks for a binary classification using sigmoid function for each unit. If the network has no hidden layer, explain why the model is equivalent to logistic regression.

2. (5 points) Consider a simple two-layer network in the lecture slides. Given m training data (x^i, y^i), $i = 1, \ldots, m$, the cost function used to training the neural networks

$$\ell(w, \alpha, \beta) = \sum_{i=1}^{m} (y^i - \sigma(w^T z^i))^2$$

where $\sigma(x) = 1/(1 + e^{-x})$ is the sigmoid function, z^i is a two-dimensional vector such that $z_1^i = \sigma(\alpha^T x^i)$, and $z_2^i = \sigma(\beta^T x^i)$. Show the that the gradient is given by

$$\frac{\partial \ell(w, \alpha, \beta)}{\partial w} = \sum_{i=1}^{m} 2(y^i - \sigma(u^i))\sigma(u^i)(1 - \sigma(u^i))z^i,$$

where $u^i = w^T z^i$. Also find the gradient of $\ell(w, \alpha, \beta)$ with respect to α and β and write down their expression.

6   Programming: Bayes and KNN classifier [25 points]

In this programming assignment, you are going to apply the Bayes Classifier to handwritten digits classification problem. Here, we use the binary 0/1 loss for binary classification, i.e., you will calculate the miss-classification rate as a performance metric.

To ease your implementation, we selected two categories from USPS dataset in usps-2cls.mat (or usps-2cls.dat, usps-2cls.csv).

1. (15 points) Your first task is implementing the classifier by assuming the covariance matrices for two classes are a diagonal matrix Σ_1, Σ_2.

   Using slides from "Classification I", assuming $P(y = 1) = P(y = -1)$ (i.e., the prior distribution for two classes are the same), using Bayes decision rule to write down the decision boundary. (Hint, it should be a quadratic decision boundary.)

   Now we will estimate the mean vector and the sample covariance matrices for two classes using the training data (hint: you can use sample mean and sample covariance vector). Report the misclassification rate (error rate) over the training set and over the testing set averaged over the 100 random train/test splits by using different value of splitting ratio p. Explain and compare the performance of each classifier.

   After implementing these methods, you should evaluate your algorithm on the given set. Repeat 100 times: split the dataset into two parts randomly, use p portion for training and the other $1 - p$ portion for testing. Let p change from 0.1, 0.2, 0.5, 0.8, 0.9.

   Please implement the algorithm **from scratch** yourself. Make sure to provide code, results (required above) together with necessary explanations to your results.

2. (10 points) Now repeat the classification again using K-nearest neighbors, for $K = 5, 10, 15, 30$. Repeat 100 times: split the dataset into two parts randomly, use p portion for training and the other $1 - p$ portion for testing. Let p change from 0.1, 0.2, 0.5, 0.8, 0.9. Report the training error and testing error for each case.

   For this part, you may use any package that you like. Make sure to provide code, results (required above) together with necessary explanations to your results.