CS4801: Principle of Machine Learning
Assignment 4

November 13, 2018

This homework consists of problems covering Clustering. A few instructions to make life easier for all of us:

- Assignment need not to be submitted.

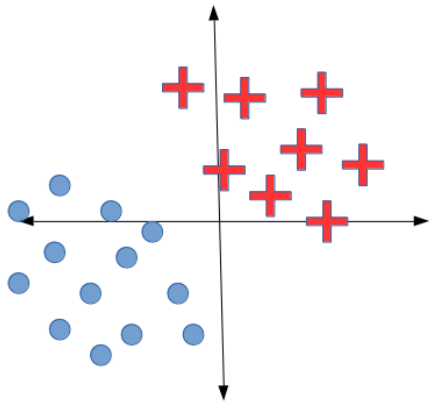
Exercise 1

(a) Consider a binary classification problem in $\mathbb{R}^2$ where each class has a normal distribution with **zero mean and unit covariance**. Suppose you train an LDA classifier on a training sample containing 1 points from each class.

i. Is it possible to apply LDA? Why?

ii. Discuss of training and test error of a classifier trained on these 2 data points?

(b) Consider the following training data points. What are the minimum possible support vector for Linear SVM and Kernel SVM with rbf kernel?



Given that we can select the same feature multiple times during the recursive partitioning of the input space, is it always possible to achieve 100% accuracy on the training data (given that we allow for trees to grow to their maximum size) when building decision trees?

Consider building a spam filter for distinguishing between genuine e-mails and unwanted spam e-mails. Assuming spam to be the positive class, which among the following would be more important to optimise Recall or precision?

Does bagging and boosting suffer from over fitting?

Exercise 2

(a) What are advantage and disadvantage of PCA and FDA(Fisher discriminative analysis)

(b) How to do non linear feature extraction ?

(c) Consider the following data set find Principal loading's and Principal components.

1 1 9
2 4 6
3 7 4
4 11 4
5 9 2

Exercise 3

(a) (2 points) Consider the following linear model with one non-random input variable X and without offset:

$$Y = Xw + \epsilon,\ E[\epsilon] = 0,\ Var(\epsilon) = \sigma^2. \quad \ldots (1)$$

Assume that we are given a set of n inputs $\{x_i\}_{i=1}^n$ (non-random) and then we draw the corresponding outputs $\{y_i\}_{i=1}^n$ i.i.d. according to (1).

Compute the bias and variance of the estimators for the *Least squares estimator* w_{LSS} and the *Ridge estimator* w_{Ridge}.

(b) How bias and variance varies with number of layers of FFNN keeping number of node in a hidden unit fixed?

Exercise 4

(a) What is backpropagation algorithm? State one drawback of it

(b) Design an neural network for OR, AND, NOT function.

Exercise 5

(a) Is the following a distribution function?

$$F(x) = e^{-\frac{1}{x}} \text{ if } x > 0,\ F(x) = 0 \text{ otherwise}$$

If so, give the corresponding density function. If not, mention why it is not a distribution function.

(b) Given N samples $x_1, \ldots, x_N$ drawn independently from a Gaussian distribution with variance σ^2 unknown mean μ. Assume that the prior distribution of the mean is also a Gaussian distribution, but with parameters mean μ_p and variance σ_p^2. Find the MAP estimate of the mean

Exercise 7

(a) Let u be a $n1$ vector, such that $u^T u = 1$. Let I be the nn identity matrix and $A = (I - kuu^T)$, where k is a real constant. If u is an eigenvector of A, with eigenvalue -1. What is the value of k?

(b) define rank of a matrix. How the rank of a data matrix is related to Principal component analysis.

Exercise 6

(a) Which clasterign algorithm will work well on following dat aste and why?