CS60050: Machine Learning

Assignment 3

Due date: Nov 8th

Q1. Suggest an algorithm for computing the bias term b in the dual SMO algorithm. Explain why this algorithm will always work.

Q2. Go back to implementation of logistic regression using SGD in the previous assignment. Implement the following variants:

1. Mini-batch SGD where you compute the stochastic gradient using a subset of examples of a specified size, say k. Vary k from from 0.1 % to 10% of the training set size.
2. Averaged SGD, where at the end of each iteration you maintain the average of last m iteration. Vary m as {1, 10, 100}
3. Minimum SGD, where at the end of each iteration you maintain the minimum of all previous iterations.

Compare the performance of the above methods for various values of m and k on the dataset Gisette available at:
https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html
Plot the test set accuracy for the parameter in each iteration.

Q3. Generate a random dataset of 10 dimensional boolean variables, approx 1500 data points. Generate the class labels using a noisy version of the following function:
$\mathrm{f(x)} = \mathrm{x}[10] + \mathrm{x}[7] + 1$, where '+' denotes xoring.
The noise probability p (probability of flipping the output bit) can be varied from 0 to 50 % in steps of 10%. Train a decision tree using the ID3 algorithm implemented in weka. Plot the test set error as a function of p.

Q4. You have two coins with unknown probabilities of heads, denoted p and q respectively.

The first coin is chosen with probability π and the second coin is chosen with probability $(1 - \pi)$. You toss the chosen coin and observe the outcome (head =1, tail =0).

1. Write a latent variable model that describes the above scenario.
2. Devise and EM algorithm to learn the parameters of the above model.
3. Generate random data from the above model, and implement the EM algorithm to learn the parameters. Report both the original parameters and learned parameters.