COMP 7745/8745: Machine Learning

Instructor: Deepak Venugopal

Spring 2016: Homework 4

Due Date: April 21, 2016 (Hard copy due before the start of class)

1. Given the following dataset, which kernel for the support vector machine would you pick and why?(15 points)

X_0	X_1	y
0	0	-1
0	1	1
1	0	1
1	1	1

2. Given a support vector machine trained on m examples, what is an upper bound on its leave-one-out cross validation error given that it has a) m support vectors b) $m/2$ support vectors (15 points)

3. Given the following dataset, we wish to apply Adaboosting using SVMs with linear kernels as our base classifier. Is this a good idea? Explain your answer. (15 points)

X_0	X_1	y
0	0	-1
0	1	-1
1	0	-1
1	1	1

4. Clustering to compress images. Here, you will implement the KMeans algorithm to compress images. You are given 2 images koala.jpg and penguins.jpg. Corresponding to these images, the RGB values for these images are given to you in data files koala.txt and penguins.txt respectively. You will implement KMeans in any language of your choice to cluster the datapoints in koala.txt and penguins.txt. You will then replace the RGB value of each datapoint with the value of its cluster-center to create a new data file which can be regarded as a compressed version of the original data. Use the showimage.java program to display the compressed image. The program takes as arguments the original jpg, the compressed data-file that you obtained after running your clustering program, and the output jpg image.

   The value of K controls the amount of compression in the image. Show the images that you obtained for K values 2,5,10 and 20. What is a good value of K for each image based on the quality of the image (i.e., at what minimum value of K is the image quality acceptable). (55 points)