Data mining - Homework 2: Support Vector Machine Algorithm

November 4, 2018

1 Problem 1: SVM algorithm for image recognition (50 points)

Consider an SVM algortihm for image recognition. Your dataset, where you will apply SVM is here: **http://yann.lecun.com/exdb/mnist/**. It is called MNIST and it consists of images of digits. Each image is encoded as a sequence of pixels which you can think of as a long feature vector (each pixel corresponds to a different dimension). The size of the feature vector is $784 = 28 \times 28$-dimensional (each picture is of size 28×28). In this exercise you will construct an SVM classifier that distinguishes between images with label 0 and images with label 9. You should extract your training set from: **train-images-idx3-ubyte.gz** (only datapoints that correspond to pictures of digits: 0 and 9). Similarly, the labels for that set should be extracted from the set: **train-labels-idx1-ubyte.gz**. The test set should be extracted from **t10k-images-idx3-ubyte.gz** and the corresponding labels from **t10k-labels-idx1-ubyte.gz**. Before applying a linear SVM that we were talking about in the class, transform each datapoint using mapping ϕ corresponding to the angular kernel (in other words, replace each datapoint with a random feature vector for that point corresponding to the angular kernel). Then train and test linear SVM for that transformed set. Explain the role of replacing original datapoints with random feature maps corresponding to the angular kernel (**5 points**). Explain how you choose the dimensionality of the random feature map, how you conduct training (in particular, which libraries for convex programming you use), report the results from training. Then explain how you conduct testing, report your final results (in particular the accuracy that you achieve on the training and test set). Is your transformed data linearly separable ? Explain how you can check it. Include the code in the solution. (**45 points**).

2 Problem 2: (20 points)

Consider a linearly separable datasets $\mathcal{X} \subseteq \mathbb{R}^d$, where datapoints have labels taken from two-element discrete datasets $\{-1, +1\}$. Propose an algorithm for finding an optimal (i.e. maximizing the minimum distance) hyperplane separating points from $\mathcal{X}$ with label $+1$ from these with label -1 that is orthogonal to a given vector $\mathbf{w} \in \mathbb{R}^d$. Prove that your algorithm constructs an optimal hyperplane (**15 points**). What is the complexity of your algorithm (**5 points**)? You can assume that a separating hyperplane orthogonal to a given vector $\mathbf{w}$ exists.