DATA 609 - Homework 6: Applications to Stats and Machine Learning

Eddie Xu

Instructions

Please submit a .qmd file along with a rendered pdf to the Brightspace page for this assignment. You may use whatever language you like within your qmd file, I recommend python, julia, or R.

Problem 1: Multi-Label Support Vector Machine (CVX Additional Exercises 6.18)

The basic SVM described in chapter 8 of the book is used for classification of data with two labels. In this problem we explore an extension of SVM that can be used to carry out classification of data with more than two labels. Our data consists of pairs:

$$(\mathbf{x}_i, y_i) \in \mathbf{R}^n \times \{1, \dots, K\},\ i = 1, \dots, m$$

where $\mathbf{x}_i$ is the feature vector and y_i is the label of the ith data point. (So the labels can take the values $1, \dots, K$.)

Our classifier will use K affine functions, $f_k(\mathbf{x}) = \mathbf{a}_k^T\mathbf{x} + \mathbf{b}_k$, $k = 1, ..., K$, which we also collect into affine function from $\mathbb{R}^n$ into $\mathbb{R}^K$ as $f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$. (Therows of A are $\mathbf{a}_k^T$.) Given the feature vector $\mathbf{x}$, our model predicts the label $\widehat{y} = \text{argmax}_k f_k(\mathbf{x})$, i.e. the predicted label is given by the index of the largest value of the f_k functions evaluated at the data point.

We assume that exact ties never occur, or if they do, an arbitrary choice can be made. Note that if a multiple of 1 is added to $\mathbf{b}$, the classifier does not change. Thus, without loss of generality, we can assume that $\mathbf{1}^T\mathbf{b} = 0$.

To correctly classify all the data examples perfectly, we would need $f_{y_i}(\mathbf{x}_i) > \max_{k \neq y_i} f_k(\mathbf{x}_i)$ for all i. This set of inequalities in a_k and b_k, are feasible if and only if the set of inequalities $f_{y_i}(\mathbf{x}_i) \geq 1 + \max_{k \neq y_i} f_k(\mathbf{x}_i)$ are feasible. This motivates the loss function: