Support Vector Machines for Classifying Handwritten Digits

Perrin Ruth Aparajithan Venkateswaran Derek Wright

James Meiss
APPM 3310, Fall 2017
University of Colorado Boulder

Description

Our project aims at classifying handwritten digits (MNIST dataset, in particular) with the help of Support Vector Machines. Classification of images has received great attention in recent times, and there are many applications. One particular application of classifying handwritten digits is automating the process of sorting mail by zip codes. Our project will make use of Support Vector Machines, a set of linear classifiers to help us classify digits.

Support Vector Machines

Support Vector Machines are a type of linear classifiers. They classify data by constructing a hyperplane that divides the data into two different classes – the positive, and the negative class. If we are given a training dataset of n points, $(x_1, y_1), \dots, (x_n, y_n)$ where $y_i = \{-1, 1\}$ where $-1, 1$ represent the two classes, then we can construct the hyperplane,

$$\vec{w} \cdot \vec{x} - b = 0 \quad (1)$$

where $\vec{w}$ is the normal to the hyperplane. If the training data is linearly separable, which is our assumption, then our desired hyperplane (1) lies halfway between these two parallel hyperplanes that form the margin of the two different classes:

$$\vec{w} \cdot \vec{x} - b = -1 \quad (2)$$

$$\vec{w} \cdot \vec{x} - b = 1 \quad (3)$$

The distance between these hyperplanes is proportional to $\frac{1}{\|w\|}$. Since we want to maximize the distance from the two different classes to hyperplane (1), we must minimize $\|w\|$. Therefore, our problem now becomes an optimization problem where we attempt to minimize the norm of the normal vector to the hyperplane that separates our data into two different classes subject to constraints given by the training dataset.

References

[1] Steve R. Gunn (1998) "*Support Vector Machines for Classification and Regression*"