Assignment 4 - Support Vector Machines

Shivakanth Thudi

2/23/2017

Part 1 - Toy SVM

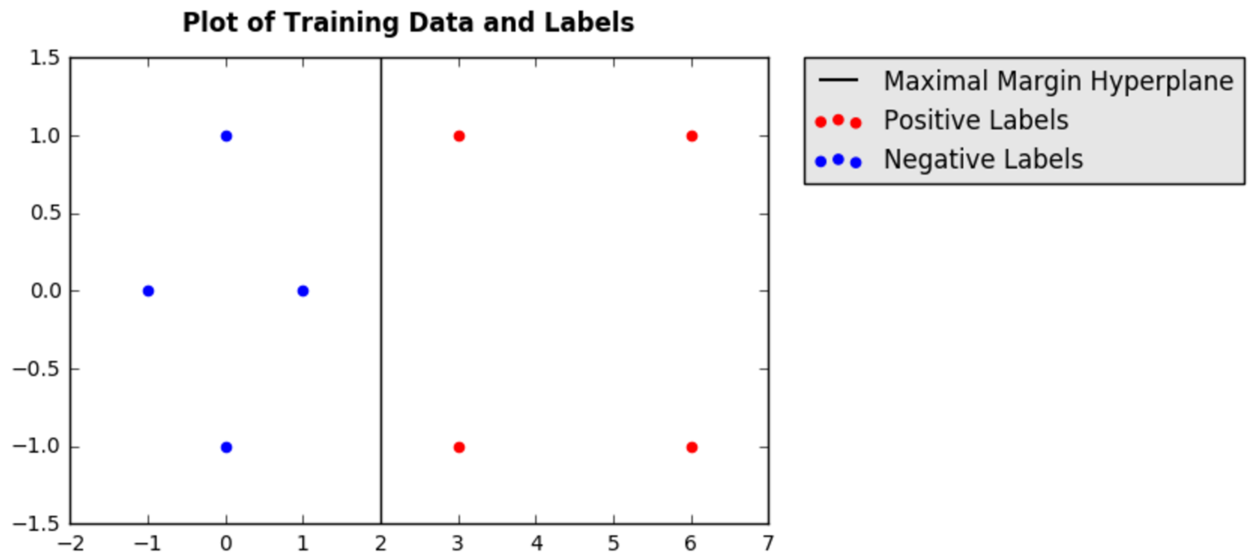

Figure 1: Scatter Plot

Toy SVM

We note this dataset is seperable and the minimum distance between any two points with opposite labels is 2.

The margin is smaller or equal to $\underline{1}$.

The line $x^1 = 2$ has margin $\underline{1}$, so this line is our SVM solution.

$\Rightarrow$ Apart from $(1,0)$, $(3,1)$, and $(3,-1)$, $(1,0)$ the other points will have Lagrangian co-efficients set to zero.

$$\Rightarrow \quad \alpha_3 = \alpha_4 = \alpha_6 = \alpha_7 = \alpha_8 = 0.$$

We have

$$w_0 + 3w_1 + 1w_2 = \underline{1}$$
$$w_0 + 3w_1 - 1w_2 = \underline{1} \qquad \Rightarrow \quad w_2 = 0$$

$$w_0 + 1w_1 + 0w_2 = -\underline{1} \qquad \Rightarrow \quad w_1 = 1$$

$$\Rightarrow \quad w_0 = -2$$

So $w_0 = -2$, $w_1 = 1$, and $w_2 = 0$

$$w_1 = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i \quad \text{and} \quad \sum y_i \alpha_i = 0$$

$$w_1 = \alpha_1 (1)(3) + \alpha_2 (1)(3) + \alpha_5 (-1)(1) \quad \ldots \quad (1)$$

$$w_2 = \alpha_1 (1)(1) + \alpha_2 (1)(-1) + \alpha_5 (-1)(0) \quad \ldots \quad (2)$$

$$\alpha_1 + \alpha_2 - \alpha_5 = 0 \quad (3)$$

$$3\alpha_1 + 3\alpha_2 - \alpha_5 = 1$$
$$\alpha_1 - \alpha_2 = 0$$
$$\alpha_1 + \alpha_2 - \alpha_5 = 0$$

$2\alpha_1 = 1 - 3\alpha_2$

$\alpha_1 =$

$\alpha_5 \; 2\alpha_1 + 3\alpha_2 = 1$

$\Rightarrow \alpha_5 = 2\alpha_1$ and $\alpha_2 = \alpha_1$

$\Rightarrow \alpha_1 = 1/4$, $\alpha_2 = \frac{1}{4}$ and $\alpha_5 = \frac{1}{2}$

(a) $\alpha_1 = \frac{1}{4}$, $\alpha_2 = \frac{1}{4}$ and $\alpha_5 = \frac{1}{2}$

$\alpha_3 = \alpha_4 = \alpha_6 = \alpha_7 = \alpha_8 = 0$

(b) $w_1 = 1$, $w_2 = 0$ and $w_0 = -2$

$w = (1, 0)$ and $w_0 = -2$

(c) The points $(3,1)$, $(3,-1)$ and $(1,0)$ are support vectors.

(d) $\hat{y}(x) = \text{Sign}\left(w_0 + \sum_{i \in S} \alpha_i y_i K(x, x_i)\right)$

$= \text{Sign}\left(w_0 + \frac{1}{4}(3x^1 + x^2) + \frac{1}{4}(3x^1 - x^2) - \frac{1}{2}(x^1 + 0x^2)\right)$

$= \text{Sign}\left(-2 + x^1\right)$

(e) $\hat{y}(x)$ for $(0.5, 0.5)$;

$\hat{y}(x) = \text{Sign}(-2 + 0.5) = \text{Sign}(-1.5)$

$= \underline{\underline{-1}}$

Assignment 4 - Support Vector Machines

Shivakanth Thudi

2/23/2017

Part 2 - Kernel SVMs

1) Quadratic and Gaussian Kernels

For quadratic kernels, we search over γ in the range of [0.01, 1] with a step size of 0.01, and over C (penalty parameter) in the range of [0.1, 10] with a step size of 0.1. We use the RandomizedSearchCV package from sklearn and look at 1000 random combinations of the parameters, rather than perform an exhaustive search over all combinations of parameters.

Kernel	Cross-Validated Accuracy	Test Set Accuracy	Gamma	C
Quadratic	0.842	0.846	0.12	0.9
Gaussian	0.841	0.845	0.03	2.6

Quadratic kernels performed slightly better than Gaussian kernels, with a test set accuracy of 0.846. The best parameters are shown in the table above.

2) kNN Classifiers

We used a k-nearest neighbors classifier to make predictions as well. We looked in the range of [1, 10] for the number of neighbors and found that 10 neighbors gave the highest cross-validated score on the training data. The testing accuracy is also shown below.

Algorithm	Cross-Validated Accuracy	Test Set Accuracy	Number of Neighbors
k-Nearest Neighbors	0.826	0.828	10

Both kernel SVMS performed better on training and testing data than k-Nearest Neighbors.

3) Results

The Quadratic and Gaussian kernel SVMs outperformed k-nearest Neighbors, with the Quadratic kernel doing slightly better than the Gaussian kernel. Support vector machines produce non-linear decision boundaries, use a subset of training points in the decision function (called support vectors), and are memory efficient.

When the features are high-dimensional, the performance of k-nearest neighbors degrades since in high-dimensional space, data points might not really be "close" to each other. Hence, the k-nearest neighbors algorithm performs worse in this situation since the features are high-dimensional.

Support vector machines, however, are more effective in high dimensional spaces than k-nearest neighbors.