Notes on Support Vector Machine

greeness

June 13, 2013

1 Notation

We use $y \in \{-1, 1\}$ to denote the class labels. We use parameters w, b and write out classifier as $g(w^T x + b)$, where b is the intercept term (or bias). From our definition of g, our classifier will directly predict either 1 or -1 [1].

2 Functional Margin

Given a training example $(x^{(i)}, y^{(i)})$, we define the function margin of (w, b) with respect to the training example

$$\hat{\gamma}^{(i)} = y^{(i)}(w^T x + b).$$

Note that if $y(i) = 1$, then for the functional margin to be large, we need $w^T x + b$ to be a large positive number. Conversely, if $y(i) =$ -1, then for the functional margin to be large, we need $w^T x + b$ to be a large negative number.

Given a training set $S = \{\}$, we define the function margin of (w, b) with respect to S to be the smallest of the functional margins of the individual training examples.

$$\hat{\gamma} = \min_{i=1,\cdots,m} \hat{\gamma}^{(i)}.$$

3 Geometric Margin

First since the decision boundary is defined by $w^T x + b = 0$, we immediately know that w is orthogonal (at 90°) to the separating hyperplane.

Consider the point at A, which represents the input $x^{(i)}$ of some training example with label $y^{(i)} = 1$. Its distance to the decision boundary, $\gamma^{(i)}$, is given by the line segment AB.

- $w/||w||$ is a unit-length vector pointing in the same direction as w;
- A represent $x^{(i)}$, $\overrightarrow{A} = x^{(i)}$
- $\overrightarrow{BA} = \gamma^{(i)} \frac{w}{||w||}$ (Think about vector length $\times$ vector direction)

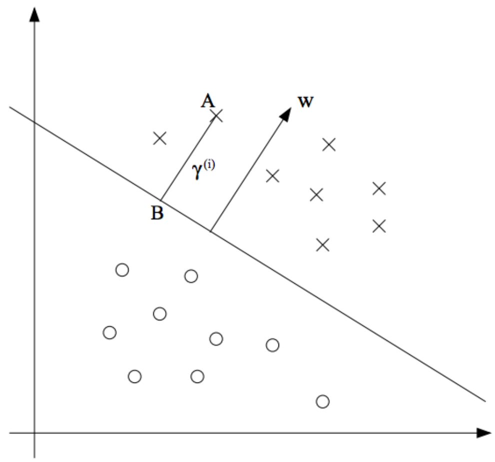

Figure 1: Geometric Margin example

- Consider $\overrightarrow{A} = \overrightarrow{B} + \overrightarrow{BA}$
- we therefore find that the point $\overrightarrow{B} = \overrightarrow{A} - \overrightarrow{BA} = x^{(i)} - \gamma^{(i)} \cdot \frac{w}{||w||}$.

Considering point B lies on the decision boundary and all points x on the decision boundary satisfy the equation $w^T x + b = 0$.

$$w^T \left(x^{(i)} - \gamma^{(i)} \cdot \frac{w}{||w||} \right) + b = 0.$$

Solving for $\gamma^{(i)}$ yields

$$\gamma^{(i)} = \frac{w^T x^{(i)} + b}{||w||} = \left(\frac{w}{||w||} \right)^T x^{(i)} + \frac{b}{||w||}.$$

The above is a solution for the case of a positive training example at A in the figure. More generally, we define the geometric margin of (w, b) with respect to a training example $(x^{(i)}, y^{(i)})$ to be

$$\gamma^{(i)} = y^{(i)} \left[\left(\frac{w}{||w||} \right)^T x^{(i)} + \frac{b}{||w||} \right].$$

Not that if $||w|| = 1$, then the functional margin equals the geometric margin. Also the geometric margin is invariant to rescaling of the parameters.

Given a training set $S = \{\}$, we define the geometric margin of (w, b) with respect to S to be the smallest of the geometric margins of the individual training examples.

$$\gamma = \min_{i=1,\cdots,m} \gamma^{(i)}.$$

References

[1] Anrew Ng, *CS229 Lecture notes - Support Vector Machine*, http://cs229.stanford.edu/notes/cs229-notes3.pdf.