[Summary]SVMs 2

Mathematics Behind Large Margin Classification

Vector Inner Product

Say we have two vectors, u and v:

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \quad v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

The **length of vector v** is denoted $|| v ||$, and it describes the line on a graph from origin (0,0) to (v_1, v_2).

The length of vector v can be calculated with $\sqrt{v_1^2 + v_2^2}$ by the Pythagorean theorem.

The **projection** of vector v onto vector u is found by taking a right angle from u to the end of v, creating a right triangle.

- p= length of projection of v onto the vector u.
- $u^T v = p \cdot || u ||$

Note that $u^T v = ||u|| \cdot ||v|| cos\theta$ where θ is the angle between u and v. Also, $p = ||v|| cos\theta$. If you substitute p for $||v|| cos\theta$, you get $u^T v = p \cdot || u ||$.

So the product $u^T v$ is equal to the length of the projection times the length of vector u.

In our example, since u and v are vectors of the same length, $u^T v = v^T u$.

$$u^T v = v^T u = p \cdot || u || \quad = u_1 v_1 + u_2 v_2$$

If the **angle** between the lines for v and u is **greater than 90 degrees**, then the projection p will be **negative**.

$$\min_{\Theta} \frac{1}{2}\sum_{j=1}^{n} \Theta_j^2$$

$$= \frac{1}{2}(\Theta_1^2 + \Theta_2^2 + \cdots + \Theta_n^2)$$

$$= \frac{1}{2}\left(\sqrt{\Theta_1^2 + \Theta_2^2 + \cdots + \Theta_n^2}\right)^2$$

$$= \frac{1}{2}||\Theta||^2$$

We can use the same rules to rewrite $\Theta^T x^i$:

$$\Theta^T x^i = p^i \cdot ||\Theta|| = \Theta_1 x_1^i + \Theta_2 x_2^i + \cdots + \Theta_n x_n^i$$

So we now have a new **optimization objective** by substituting $p^i \cdot ||\Theta||$ in for $\Theta^T x^i$

- if y=1, we want $p^i \cdot ||\Theta|| \geq 1$
- if y=0, we want $p^i \cdot ||\Theta|| < -1$

The reason this causes a "large margin" is because: the vector for Θ is perpendicular to the decision boundary. In order for our optimization objective (above) to hold true, we need the absolute value of our projections p^i to be as large as possible.

If $\Theta_0 = 0$, then all our decision boundaries will intersect (0,0). If $\Theta_0 \neq 0$, the support vector machine will still find a large margin for the decision boundary.