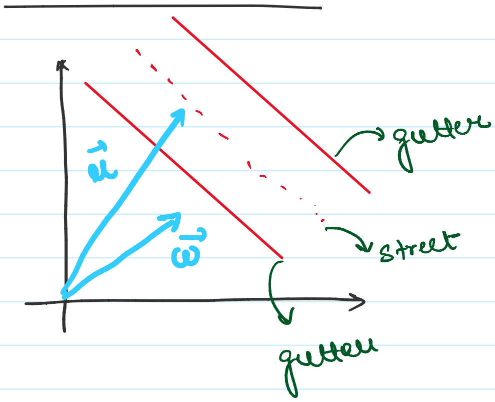
Support Vector Machines (Pepcoding)

18 June 2024 01:14

Maths behind SVM



We want to maximize the gap

We find projection of $\vec{u}$ on $\vec{w}$

$\vec{w} \cdot \vec{u} \geq c$, c is constant

Now let $c = -b$, for mathematical ease

$\therefore$ $\vec{w} \cdot \vec{u} + b \geq 0$ → Decision Rule

now depending on the value we classify as +ve / -ve sample

$\vec{w}$ = ⊥r to the gutter

$\vec{u}$ = unknown vector

x_+ → +ve sample x_- → -ve sample

Eq^{n}s for +ve & -ve sample respectively

$\vec{w} \cdot x_+ + b \geq 1$ → +ve sample → above street

$\vec{w} \cdot x_- + b \leq -1$ → -ve sample → below the street

The separation b/w samples range from -1 to +1

For mathematical ease, y is introduced

$y_i = 1$ → +ve samples $y_i = -1$ → -ve samples

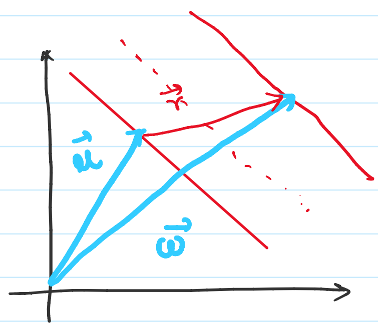
Multiplying $y_i = 1$ and $y_i = -1$ respectively

$$\left.\begin{aligned} y_i(\vec{w}x_+ + b) &\geq 1 \\ y_i(\vec{w}x_- + b) &\geq 1 \end{aligned}\right\} \text{generalized eq}^n \Rightarrow y_i(\vec{w}x_i + b) - 1 \geq 0$$

For points on the gutter ⇒ $y_i(\vec{w}x_i + b) = 0$

unit vector

the



Width of Street $= x_+ - x_- \cdot \frac{\vec{w}}{\|w\|}$ ← unit vector ⊥r to the street

To maximize the width of the street

$$y_i(\vec{w}x_i + b) - 1 = 0$$

$y_i = 1$ for +ve samples $y_i = -1$ for -ve samples

When $y_i = 1$ $\vec{w}x_+ = 1 - b$

When $y_i = -1$ $\vec{w}x_- = 1 + b$

$$\text{Width of Street} \Rightarrow (x_+ - x_-)\cdot\frac{\vec{w}}{\|w\|} \Rightarrow \frac{\vec{w}x_+ - \vec{w}x_-}{\|w\|}$$

$$\Rightarrow \frac{1 + \cancel{b} + 1 - \cancel{b}}{\|\vec{w}\|}$$

$$\rightarrow \frac{2}{\|\vec{w}\|} \text{ maximize}$$

$\therefore \|\vec{w}\|$ minimize

$\therefore \Rightarrow \frac{1}{2}\|\vec{w}\|^2$ is minimized

To find the extreme values with multiple constraints

Langrangeh multiplier eqⁿs

$$L = \frac{1}{2}\|\vec{w}\|^2 - \sum_i \alpha_i \left[y_i(\vec{w}\vec{x}_i + b) - 1 \right]$$

Differentiating wrt $\vec{\omega}$

$$\frac{\partial L}{\partial \vec{\omega}} = \vec{\omega} - \sum_i \alpha_i (y_i x_i) = 0$$

$$\Rightarrow \vec{\omega} = \sum_i \alpha_i (y_i x_i) \rightarrow \text{linear sum of samples}$$

Differentiating wrt b

$$\frac{\partial L}{\partial b} = -\sum_i \alpha_i y_i = 0$$

$\because$ Summation of constant = 0

Putting value of $\vec{\omega}$

$$L = \frac{1}{2} \sum \alpha_i (y_i x_i) \sum \alpha_i (y_i x_i) - \sum \alpha_i [y_i (\sum \alpha_i (y_i x_i) x_i + b - 1)]$$

i = +ve

j = -ve

$$L = \frac{1}{2} \sum_i (\alpha_i y_i x_i) \cancel{\sum_j (\alpha_j y_j x_j)} - \cancel{\sum_i (\alpha_i y_i x_i)} - \sum_j \alpha_j y_j x_j - \sum \alpha_i y_i b + \sum \alpha_i$$

$$= \sum \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i x_j$$

Putting value of $\vec{\omega}$ in Decision rule

$$\alpha_i y_i x_i + \vec{u} + b \geq 0$$