SVMs

Same solver can be used for the linearly seperable and the not linearly seperable.

$$\arg\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

$$\text{s.t.} \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

This called quadratic program, and there are solvers for it.

This the constrained when its linearly seperable.

$$0 \le \alpha_i \le C$$

This the constrain when its not linearly seperable.

We have nearly the same optimization task as before however, $0 \le \alpha_i \le C$. This is a quadratic program that can be solve in a couple lines of code.

Other points

- Many α_i will be $\emptyset$. The non-zero α_i are nown as the suport vectors

$$W = \sum_{i \in SV} \alpha_i y_i x_i$$

C is reguliation term

C is dataset dependent

- If C is very large the margin will be small

It means we have very few support vector

- " " small " " be large

- $$\hat{y} = \underbrace{\left(\sum_{i \in SV} \alpha_i y_i x_i\right)}_{W}^{T} x + b = \sum_{i \in SV} \alpha_i y_i K(x_i, x) + b$$

$$= W^T x + b$$

$$\boxed{K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)}$$

[kernel]

$$K: \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$$

The phi function transform x to higher dimension space

$x \in \mathbb{R}^2$

[Motivating the kernel trick]

$x = (x_1, x_2)^T$

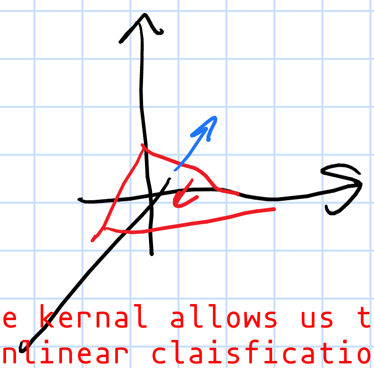
$$\Phi(x)^T \Phi(x) = \begin{bmatrix} x_1^2 \\ \sqrt{2} x_1 x_2 \\ x_2^2 \end{bmatrix}^T \begin{bmatrix} x_1^2 \\ \sqrt{2} x_1 x_2 \\ x_2^2 \end{bmatrix}$$

$$= x_1^2 x_1^2 + 2 x_1^2 x_2^2 + x_2^2 x_2^2$$

$$= (x_1^2 + x_2^2)^2$$

$$= \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right)^2$$

$$= \underline{(x^T x)}^2 = K(x, x)$$



The kernal allows us to solve nonlinear claisfication tasks.

That is equivalint to taking the dot product in high dim space 3d

$$K_1(x, z) = (x^T z)^p$$

$$K_2(x, z) = (\alpha x^T z + \beta)^p = \Phi(x)^T \Phi(z)$$

This a higher demitional product.

$$K(x, z) = \exp\left(-\gamma \|x - z\|_2^2\right) \qquad \gamma > 0$$

$$= \Phi(x)^T \Phi(z)$$

$$= \sum_{n=0}^{\infty} \frac{(-\gamma \|x - z\|_2^2)^n}{n!} = \sum_{n=0}^{\infty} \tilde{x}_n \tilde{z}_n$$