CS4780 Lec 24 – Kernel SVM

Linear

$$\min_{w} \sum_{i=1}^{n} (w^T x_i - y_i)^2$$

$X = [x_1 \dots x_n]$

$Y = [y_1 \dots y_n]$

$$\min_{w} (X^T w - Y^T)^2$$

$[X^T X]_{ij} = \vec{x}_i^T \vec{x}_j$

with $w = X\alpha = \sum_{i=1}^{n} \vec{x}_i \alpha_i$

$$\min_{w} (\underbrace{X^T X}_{K} \alpha - Y^T)^2$$

$$\min_{\alpha} (K\alpha - Y^T)^2$$

complete square

$$\alpha^T K^T K \alpha - 2 y K \alpha + Y^T Y$$

minimize

$$\frac{\partial}{\partial \alpha} \Rightarrow 2KK\alpha - 2KY^T = 0$$

$$KK\alpha = KY^T \quad] \times K^{-1}$$

$$K\alpha = Y^T \quad] \times K^{-1}$$

$$\boxed{\alpha = K^{-1} Y^T}$$ — No feature importance

Test point:

$$h(z) = w^T z$$
$$= \alpha^T X^T z$$
$$= \underbrace{\alpha^T K_*}$$

Back to linear classifier during test time back to n dimension (dim. of α)

$X^T z$ = inner product to test point z with every single point in training data

$K_* = z^T x_i$
$= K(z, x_i)$

Parametric algorithm – Fixed size of parameter learned. If more data, you don't learn more parameters.

Non-parametric algo – Model size grows as training data grows
Eg. KNN

SVM:

$$\min_{w,b} \; w^T w + C \sum_{i=1}^{n} \xi_i$$

$$\text{s.t. } \forall i \quad y_i (w^T x_i + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0$$

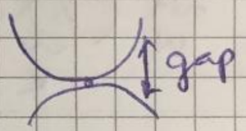
Kernelized version:

Every convex has a dual

minimum has exactly same pt. as maximum in dual space

primal - dual gap

minimize to zero



$$\min_{\alpha_1 \dots \alpha_n} \frac{1}{2} \sum \alpha_i \alpha_j y_i y_j \underbrace{K_{ij}}_{\text{innerproduct}} - \sum \alpha_i$$

$$\text{s.t. } 0 \leq \alpha_i \leq C$$

$$\sum_{i=1}^{n} \alpha_i \alpha_j = 0$$

only those points that lie on margin have non zero α

Solve for w

$$w = \sum_{i=1}^{n} \alpha_i y_i x_i$$

hypothesis test

$$h(z) = w^T z + b$$

$$= \sum_{i=1}^{n} \alpha_i y_i k(x_i, z) + b$$