Assignment

(4) Kernel SVM with TRT
(Lets consider binary SVM)

$$R(y, f(x)) = \frac{1}{n}\sum_{i=1}^{n}\left[1 - y_i f(x_i)\right]_+ + ~~\max(x, 0)~~$$

$\Rightarrow$ Consider only positive losses

~~$f(x_i)$~~

Replacing $f(\cdot) = \sum_{j=1}^{n} \alpha_j K(\cdot, x_j)$

$$R(y, f(x)) = \frac{1}{n}\sum_{i=1}^{n}\left(1 - y_i \sum_{j=1}^{n} \alpha_j K(x_i, x_j)\right) +$$

$$R(y, f(x)) = 1 - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i \alpha_j K(x_i, x_j) +$$

$$J(\alpha) = 1 - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i \alpha_j K(x_i, x_j) +$$

in case of an l_2 regularized SVM

$$R(y_i, f(x_j)) = 1 - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i \alpha_j K(x_i, x_j)$$

$$+ \lambda \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j K(x_i, x_j)$$

$$= 1 - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i \alpha_j K_{ij} + \lambda \alpha^T K \alpha$$

$$J(\alpha) = 1 - \frac{y^T K \alpha}{n} + \lambda \alpha^T K \alpha$$

$$-\frac{K^T Y}{n} + \lambda\left[2K\alpha\right] = 0.$$

$K^T = K$

$$\Rightarrow 2\lambda K\alpha = \frac{KY}{n}$$

$$\boxed{\alpha = \frac{Y}{2\lambda n}}$$

Resubstituting α back in the equation gives solution similar to solving duals of SVM

(1) Assuming

$$\tilde{\phi}_i = \phi_i - \frac{1}{n}\sum_{i=1}^{n} \phi_i$$

$$\tilde{K} = \frac{1}{n}\sum_{i=1}^{n} \tilde{\phi}_i \, \tilde{\phi}_i^T$$

$$\tilde{K} = \frac{1}{n}\sum_{i=1}^{n}\left[\phi_i - \frac{1}{n}\sum_j \phi_j\right]\left[\phi_i - \frac{1}{n}\sum_j \phi_j\right]^T$$

$$\tilde{K} = \frac{1}{n}\sum_{i=1}^{n}\left[\phi_i\phi_i^T - \frac{1}{n}\phi_i\sum_j \phi_j^T - \frac{1}{n}\phi_i\sum_j \phi_j^T + \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\phi_j\phi_k^T\right]$$

$$\therefore \tilde{K} = \frac{1}{n}\sum_{i=1}^{n}\phi_i\phi_i^T - \frac{2}{n^2}\sum_{i=1}^{n}\phi_i\sum_{j=1}^{n}\phi_j^T + \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\phi_j\phi_k$$

in matrices form this can be written as

$$\boxed{\tilde{K} = K - 2\,\mathbb{1}_{1/n}K + \mathbb{1}_{1/n}K\,\mathbb{1}_{1/n}.}$$

② The Representer Theorem (2001):-

Let K be a kernel of $\mathcal{X}$ and let $\mathcal{H}$ be the RKHS associated with it. Fix $x_1 \ldots x_n \in \mathcal{X}$ and

Consider the optimization

$$f^* = \underset{f \in \mathcal{H}}{\operatorname{argmin}} \; R\left(f(x_1) \ldots f(x_n)\right) + g\left(\|f\|_{\mathcal{H}}^2\right)$$

where $g(\cdot)$ is non increasing and R depends on x_i only through $f(x_i)$.

If f^* has a solution then its of the form

$$f(\cdot) = \sum_i \alpha_i K(\cdot, x_i)$$

If $g(\cdot)$ is an increasing function then every solution of f^* is of the same form

Proof:

Let $f = f_{\parallel} + f_{\perp}$

$f_{\parallel}$ = projection of f onto the span of kernels

$f_{\perp}$ = orthogonal to kernel functions

$$\|(f_{\parallel})\|^2 + \|f_{\perp}\|^2 = \|f\|^2$$

$$\Rightarrow \|f\|^2 \geq \|f_{\parallel}\|^2$$

$$\therefore g(\|f\|)^2 \geq g(\|f_{\parallel}\|)^2$$

for most optimal case during optimization

$$g(\|f\|^2) = g(\|f_{\parallel}\|^2)$$

$$\Rightarrow \underline{\underline{f_{\perp} = 0}}$$

as a result we can express

$$f(\cdot) = \sum_{i=1}^{n} \alpha_j K(\cdot, x_j)$$

where K are kernel functions

③ TRT for ridge regression

$$f^* = \underset{f \in \mathcal{H}}{\operatorname{argmin}} \ \frac{1}{n} \sum_{i=1}^{n} \left(y_i - f(x_i) \right)^2 + \lambda \| f \|^2$$

with TRT we get

$$f^* = \underset{f \in \mathcal{H}}{\operatorname{argmin}} \ \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{n} \alpha_j K(\cdot, x_j) \right)^2$$

$$+ \lambda \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j K(x_i, x_j)$$

In matrix form we have

$$f^* = \underset{\alpha}{\operatorname{argmin}} \ \underbrace{\frac{1}{n} \| Y - K\alpha \|^2 + \lambda \alpha^T K \alpha}_{J(\alpha)}$$

Assuming K is invertible, $J(\alpha)$ is convex in α

Setting $\nabla J(\alpha)$ to zero we have

$$\nabla\left((y-K\alpha)^T (y-K\alpha) + \lambda \alpha^T K \alpha \right) = 0$$

$$-\frac{1}{n}\left[y - K\alpha \right] + \lambda\alpha = 0$$

$$\Rightarrow \boxed{\alpha = \left[K + n\lambda I \right]^{-1} Y}$$