Date: __/__/____

Machine Learning - B555
Written Assignment 4

Ans 1. To prove: $\frac{1}{(\gamma^2)} = \sum_i \alpha_i^*$

γ_i = distance from line of margin

Objective = $\max_{w, w_0} \min \gamma_i$ Minimum value of γ_i

We know, $\gamma_i = t_i (w^T \phi(x_i) + w_0) \geq \gamma$

Divide eq^n by γ

$\therefore \quad t_i \times \frac{1}{\gamma} (w^T \phi(x_i) + w_0) \geq 1$ — (1)

Consider $V = \frac{w}{\gamma}$,

$\therefore \quad \|V\|^2 = \frac{\|w\|^2}{\|\gamma\|^2} \qquad \|w\| = 1$

$\|V\|^2 = \frac{1}{\gamma^2}$ — (2) This changes the objective to $\min \frac{1}{2}\|V\|^2$

Put in (1)

$t_i (V^T \phi(x_i) + V_0) \geq 1$ This is the constraint to

This objective along with the constraint can be solved using the Lagrangian.

$\mathcal{L}(p, \alpha, \beta) = f(p) + \sum_j \beta_j h_j(p) + \sum_k \alpha_k g_k(p)$

#learnthesmarterway

Date: __/__/____

Here $g_k(p) = g_i(v) = 1 - t_i(v^T \phi(x_i) + v_0) \leq 0$

$f(p) = f(v) = \frac{1}{2} v^T v$

& $h_j(p) = 0$

Putting all this in the Lagrangian

Min $f(p)$ subject to $g_k(p) \leq 0$ & $h_j(p) = 0$

~~$L($~~

$$L(v, \alpha, v_0) = \frac{1}{2} v^T v + \sum_i \alpha_i \left(1 - t_i(v^T \phi(x_i) + v_0)\right)$$

$$\frac{dL}{dv} = v - \sum_i \alpha_i t_i \phi(x_i) = 0$$

$\therefore \boxed{v = \sum_i \alpha_i t_i \phi(x_i)} \rightarrow$ Same form as dual perceptron for optimal sol^n.

$$\frac{\partial L}{\partial v_0} = \boxed{\sum \alpha_i t_i = 0.}$$

Cannot solve this directly

So, we use dual optimisation

$$\theta(\alpha) = \min L(v, \alpha, v_0)$$

s.T $g_i(v) \leq 0$

Then Maximise $\theta(\alpha)$

such that $\alpha_i \geq 0$.

Now $$\theta(\alpha) = \frac{1}{2} \sum_i \sum_k \alpha_i \alpha_k t_i t_k \phi(x_i) \phi(x_k)^T$$

$$+ \sum_i \alpha_i - \cancel{\sum \alpha_i t_i v_0}$$

$$- \sum_i \alpha_i t_i \left[\sum_k \alpha_k t_k \phi(x_k)^T \cdot \phi(x_i) \right]$$

$$\theta(\alpha) = \sum \alpha_i - \frac{1}{2} \sum_i \sum_k \alpha_i \alpha_k t_i t_k \boxed{\phi(x_i) \phi(x_k)^T} \rightarrow K$$

$$\theta(\alpha) = \sum_i \alpha_i - \frac{1}{2}\sum_i \sum_k \alpha_i \alpha_k t_i t_k K(x_i, x_k)$$

Such that $\alpha_i \geq 0$ & $\sum \alpha_i t_i = 0$

$$\theta(\alpha) = \sum \alpha_i - \frac{1}{2}\|V\|^2 \quad \text{— (A)}$$

Now $\theta(\alpha) \overset{\max}{=} \mathcal{L}(V, \alpha, V_0)$

$f(p) \geq \mathcal{L}(p, \alpha, \beta)$ based on the constraints

So $\max^m$ value of $\mathcal{L}(p, \alpha, \beta) = f(p)$

Put in (A)

$$\frac{1}{2}\|V\|^2 = \sum \alpha_i - \frac{1}{2}\|V\|^2$$

$$\|V\|^2 = \sum \alpha_i$$

or $\boxed{\frac{1}{r^2} = \sum \alpha_i}$ where $\|V\|^2 = \frac{1}{r^2}$ (derived earlier)

Ans. 2) Objective: $\text{Min } \frac{1}{2}V^T V + \frac{1}{2}C\sum_i \xi_i^2$

S.t $t_i(V^T\phi(x_i) + V_0) \geq 1 - \xi_i$

$g_i(V) = 1 - \xi_i - t_i(V^T\phi(x_i) + V_0) \leq 0$

where $g_i(V) \leq 0$

$$\mathcal{L}(V, \alpha, V_0, \xi) = \frac{1}{2}V^T V + \frac{1}{2}C\sum_i \xi_i^2 + \sum_i \alpha_i(1 - \xi_i - t_i(V^T\phi(x_i) + V_0))$$

Here $\alpha_i \geq 0$, $g_i(V) \leq 0$

$$\frac{\partial \mathcal{L}}{\partial V} = V - \sum_i \alpha_i t_i \phi(x_i) = 0$$

$$\boxed{V = \sum_i \alpha_i t_i \phi(x_i)}$$

Date: __/__/____

$$\frac{\partial L}{\partial v_0} = \boxed{\sum \alpha_i t_i = 0.}$$

$$\frac{\partial L}{\partial z_i} = C\sum_i z_i - \sum \alpha_i - \cancel{\sum_i \alpha_i z_i} = 0$$

$$\boxed{\cancel{\alpha_i = \frac{\sum_i \alpha_i - C\sum z_i}{2}}} \qquad \sum \alpha_i = C\sum_i z_i$$

$$\theta(\alpha, z_i) = \frac{1}{2}\sum_i\sum_k \alpha_i\alpha_k t_i t_k \phi(x_i)^T \phi(x_k) + \frac{1}{2}C\sum_i z_i^2$$

$$+ \sum_i \alpha_i - \sum_i \alpha_i z_i - \cancel{\sum \alpha_i t_i v_0}$$

$$- \sum_i \alpha_i t_i \left(\sum_k \alpha_k t_k \phi(x_k)^T \phi(x_i)\right)$$

$$\cancel{\theta \sum_i \alpha_i z_i^2 - \sum_i}$$

$$-\sum_i \cancel{\frac{\alpha_i z_i^2}{2}}^{2} + \cancel{\sum_i \frac{C\alpha_i^3}{2}}$$

$$= -\frac{1}{2}\sum_i\sum_k \alpha_i\alpha_k t_i t_k \phi(x_i)^T \phi(x_k) + \sum_i \alpha_i$$

$$- \sum_i \alpha_i z_i + \frac{1}{2}\sum_i \alpha_i z_i$$

$$= \sum_i \alpha_i - \frac{1}{2}\sum_i \alpha_i z_i - \frac{1}{2}\sum_i\sum_k \alpha_i\alpha_k t_i t_k \underbrace{\phi(x_i)^T \phi(x_k)}_{k(x_i, x_k)}$$

Dual form:

$$\theta(\alpha, z_i) = \sum_i \alpha_i \left(1 - \frac{z_i}{2}\right) - \frac{1}{2}\sum_i\sum_k \alpha_i\alpha_k t_i t_k k(x_i, x_k)$$

$$= \sum \alpha_i \left(1 - \frac{z_i}{2}\right) - \frac{1}{2}\|v\|^2$$

#learnthesmarterway

Date: __/__/____

$$= \sum_i \alpha_i\left(1 - \frac{3_i}{2}\right) - \frac{1}{2} V^T V$$

where $\alpha_i \geq 0$ & $\sum \alpha_i t_i = 0$.

We also proved that it's a kernel method by expressing $\phi(X_i)^T \times \phi(X_k) = k(X_i, X_k)$

Ans.3. Objective: $\sum_i \ell(a_i, t_i) + R(w^T w)$

$$\text{Here } a_i = w^T \phi(x_i) \quad -(1)$$

Representer Theorem:

$$\frac{\partial E}{\partial W} = \sum \ell'(a_i, t_i) \times \phi(X_i) + 2R' W = 0$$

$$\therefore w = -\frac{\sum \ell'(a_i, t_i) \times \phi(X_i)}{2R}$$

Here w is the optimal sol^n which can also be written as $\boxed{w = \sum_i \alpha_i \phi(X_i)}$ or $w = \Phi^T \alpha \quad -(2)$

$$\text{Where } \alpha_i = \frac{-1}{2R} \ell'(a_i, t_i)$$

$$\text{or } \alpha = \frac{-1}{2R} \ell'(\Phi w, t) \quad [\text{From } (1)]$$

$$\text{or } \alpha = \frac{-1}{2R} \ell'(\Phi \times \Phi^T \alpha, t) \quad [\text{From } (2)]$$

$$\text{or } \alpha = \frac{-1}{2R} \ell'(\Phi \Phi^T \alpha, t)$$

→ Kernel form

$$\text{Prediction} = w^T \phi(z)$$

$$= \sum \alpha_i \phi(X_i)^T \times \phi(z)$$

$$\boxed{\text{Prediction} = \sum \alpha_i K(X_i, z)}$$

Ans 04 We know that the Marginal likelihood is given by Gaussian Process

$L = N(t|0, C)$

where $C = \theta_0 K(x_n, x_m) + B^{-1}I$

↓Hyperparameter of prior

Now $C_N = \theta_0 K(X_n, X_m)$

$\frac{\partial C_N}{\partial \theta} = K(X_n, X_m)$ — (1)

Also $C_N^{-1} = \frac{1}{\theta_0} K^{-1}(x_n, x_m)$ — (2)

So We know $P(t|\theta) = \left(\frac{C_N^{-1}}{2\pi}\right)^{\frac{1}{2}} \times \exp\left\{\frac{-C_N^{-1}}{2}(t-0)^2\right\}$

$\therefore \log P(t|\theta) = -\frac{1}{2}\log|C_N| - \frac{N}{2}\log 2\pi - \frac{1}{2}t^T C_N^{-1} t$

$\frac{\partial}{\partial \theta_0} \log p(t|\theta_0) = -\frac{1}{2}\mathrm{Tr}\left(C_N^{-1}\frac{\partial C_N}{\partial \theta_i}\right) + \frac{1}{2}t^T C_N^{-1}\frac{\partial C_N}{\partial \theta_i} C_N^{-1} t$

[Using Eq^ns 6.69 & 6.70 from ToB]

$= -\frac{1}{2}\mathrm{Tr}\left(\frac{1}{\theta_0}\underbrace{K^{-1}(x_n, x_m) \times K(x_n, x_m)}_{I}\right)$

$+ \frac{1}{2}t^T \times \frac{1}{\theta_0}\underbrace{K^{-1}(x_n, x_m) \times K(x_n, x_m)}_{I} \times \frac{1}{\theta_0} K^{-1}(x_n, x_m) \times t$

[Using (1) & (2)]

$= -\frac{1}{2}\mathrm{Tr}\left(\frac{1}{\theta_0}\cancel{K}\right) + \frac{1}{2}t^T \times \frac{1}{\theta_0^2} \times K^{-1}(x_n, x_m) \times t$

$= -\frac{1}{2}\sum_{1}^{N}\frac{1}{\theta_0} + \frac{1}{2\theta_0^2} t^T K^{-1} t$

Date: __/__/____

$$\Rightarrow \frac{N}{2\theta_0} = \frac{1}{2\theta_0^2} t^T K^{-1}(x_n, x_m)\, t$$

$$\Rightarrow \boxed{\theta_0 = \frac{1}{N} t^T K^{-1}(x_n, x_m)\, t}$$