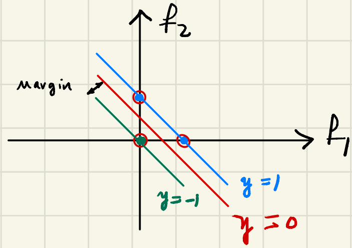
Tutorial 3



Q1.1 All data points will be SV.
Hence: $a_1, a_2, a_3 > 0$

Q1.2 Pri ·l Lagrangian:

$$\mathcal{L} = \frac{1}{2} w^T w + \sum_{n=1}^{3} a_n \left(1 - t_n (w^T x_n + b)\right)$$

$$= \frac{1}{2} w^T w - w^T \sum_{n=1}^{3} a_n t_n x_n - b \sum_{n=1}^{3} a_n t_n + \sum_{n=1}^{3} a_n$$

Optimality:

$$\nabla_w \mathcal{L} = w - \sum_{n=1}^{3} a_n t_n x_n = 0 \Rightarrow w = \sum_{m} a_m t_m x_n$$

$$w = -a_1 \begin{pmatrix} 0 \\ 0 \end{pmatrix} + a_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + a_3 \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Replacing (*) in $y(x)$:

$$y(x) = \left(\sum_{m=1}^{3} a_m t_m x_m\right)^T x + b$$

$$= \sum_{m=1}^{3} a_m t_m x_m^T x + b$$

$$= \sum_{m=1} a_m t_m \underbrace{K(x_m, x)}_{\text{Kernel of } x_m \text{ and } x} + b$$

In reality most of a_n's $= 0$, which means the pre. iction is sparse.

Q2.1

$$k_1(x,y) = \phi(x)^T \phi(y) \leftarrow \text{definition of valid kernel}$$

$$k_2(x,y) = \psi(x)^T \psi(y)$$

$$k_1(x,y) + k_2(x,y) = \phi^T(x)\phi(y) + \psi^T(x)\psi(y)$$

$$= \underbrace{(\phi(x) \mid \psi(x))}_{\lambda(x)}{}^T \underbrace{\begin{pmatrix} \phi(y) \\ \hline \psi(y) \end{pmatrix}}_{\lambda(y)}$$

$$= \lambda^T(x)\lambda(y) \checkmark$$

Q2.2

$$x^T(\lambda_i e_i e_i^T) y = (\sqrt{\lambda_i}\, x^T e_i)\cdot(\sqrt{\lambda_i}\, e_i^T y)$$

$$= \underbrace{(\sqrt{\lambda_i}\, x^T e_i)}_{\phi_i(x)} \underbrace{(\sqrt{\lambda_i}\, y^T e_i)}_{\phi_i(y)} \checkmark$$

*imp $\lambda_i \geq 0$

Q2.3 This can be derived using above results.

Q2.4

$$(x^T y)^n = (x_1 y_1 + x_2 y_2)^n$$

$$= (x_1 y_1 + x_2 y_2) \times \dots \times (x_1 y_1 + x_2 y_2) \leftarrow n\text{-factors}$$

$$= \sum_{i=1}^{n} c_i (x_1 y_1)^i (x_2 y_2)^{n-i}, \quad c > 0$$

$$= \sum_{i=1}^{n} \underbrace{(\sqrt{c_i}\, x_1^i x_2^{n-i})}_{\phi_i(x)} \underbrace{(\sqrt{c_i}\, y_1^i y_2^{n-i})}_{\phi_i(y)}$$

$$= \phi^T(x)\phi(y) \checkmark$$

Q2.5 this can be derived using Q2.1 and Q2.4

Q2.6

Recall: $e^{x} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$

Hence: $e^{x^T y} = 1 + x^T y + \frac{(x^T y)^2}{2!} + \cdots$

Each of terms above is a valid Kernel (Q 2.4)
Their summation is also valid. ✓

Q3.

SOLUTION:

$H(S) = -4/8 \log_2(4/8) - 4/8 \log_2(4/8) = 1$

$H_{f_2} = -5/8[4/5 \log_2(4/5) + 1/5 \log_2(1/5)] - 3/8[3/3 \log_2(3/3) + 0 \log_2(0/3)] = 0.4512$

$H_{f_1} = -4/8[2/4 \log_2(2/4) + 2/4 \log_2(2/4)] - 4/8[2/4 \log_2(2/4) + 2/4 \log_2(2/4)] = 1$

$H(S) - H_{f_2} = 0.5488$

$H(S) - H_{f_1} = 0$

Hence, ID3 will split on the second feature first.