103062512 Assignment 3

1. $w^* = (\lambda I_d + X^T X)^{-1} X^T r$

$$(\lambda I_d + X^T X)^{-1} X^T = (\lambda I_d + X^T I_N X)^{-1} X^T$$

$$= \left[\frac{1}{\lambda} I_d - \frac{1}{\lambda^2} X^T (I_N + \frac{1}{\lambda} X X^T)^{-1} X\right] X^T$$

$$= \left[\frac{1}{\lambda} (I_d - X^T (\lambda I_N + X X^T)^{-1} X)\right] X^T$$

$$= \frac{1}{\lambda} (X^T - X^T (\lambda I_N + X X^T)^{-1} X X^T)$$

$$= \frac{1}{\lambda} (X^T (\lambda I_N + X X^T)^{-1} (\lambda I_N + X X^T) - X^T (\lambda I_N + X X^T)^{-1} X X^T)$$

$$= \frac{1}{\lambda} \left[X^T (\lambda I_N + X X^T)^{-1}\right] (\lambda I_N + X X^T - X X^T)$$

$$= X^T (\lambda I_N + X X^T)$$

$$\therefore w^* = X^T (X X^T + \lambda I_N)^{-1} r$$

2. (a)

$$\langle g, f \rangle = \sum_{j=1}^{m} \sum_{i=1}^{n} \beta^{(j)} \alpha^{(i)} k(y^{(j)}, x^{(i)})$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha^{(i)} \beta^{(j)} k(x^{(i)}, y^{(j)})$$

$$= \langle f, g \rangle$$

2. (b)

$$\text{Let } h = \sum_{k=1}^{p} r^{(k)} k(z^{(k)}, \cdot)$$

$$af + bg = a \sum_{i=1}^{n} \alpha^{(i)} k(x^{(i)}, \cdot) + b \sum_{j=1}^{m} \beta^{(j)} k(y^{(j)}, \cdot)$$

$$\langle af + bg, h \rangle = a \sum_{i=1}^{n} \sum_{k=1}^{p} \alpha^{(i)} r^{(k)} k(x^{(i)}, z^{(k)}) + b \sum_{j=1}^{m} \sum_{k=1}^{p} \beta^{(j)} r^{(k)} k(y^{(j)}, z^{(k)})$$

$$= \alpha \langle f, h \rangle + b \langle g, h \rangle \quad \#$$

2. (c) $\langle f, f \rangle = \sum_{i=1}^{n} \alpha^{(i)^2} k(x^{(i)}, x^{(i)}) \geq 0$ since $\alpha^{(i)^2} \geq 0$ and $k(x^{(i)}, x^{(i)}) \geq 0$

and for $k(x^{(i)}, \cdot) = k(0, \cdot)$

$$k(x^{(i)}, x^{(i)}) = 0$$

$\therefore$ equality holds when $f(\cdot) = 0(\cdot)$ #

3. By the formula for distance of hyperplanes

$$\text{dist}(y_1, y_2) = \frac{|b_1 - b_2|}{\|W\|_2}$$

where $y_1 = W^T x + b_1$

$y_2 = W^T x + b_2$

$\therefore$ Distance between $\{x : W^T x - b = 1\}$ and $\{x : W^T x - b = -1\}$

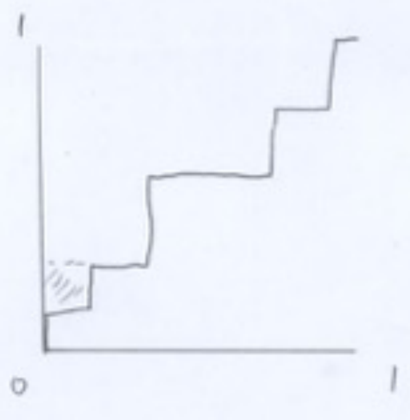
is $\dfrac{|(-b-1) - (-b+1)|}{\|W\|_2} = \dfrac{2}{\|W\|_2}$ #

6

If we partition the AUC vertically,
we will find that height of each partition represents the probability that a positive instance is chosen.

And the width (shaded area) means given a positive instance is chosen, the probability of choosing a negative instance ranked before it.

So the rest of the area (AUC part) means



the probability of choosing a negative instance after a positive instance.

Thus we found that the AUC is equal to the probability of ranking a positive instance higher than a negative one.