Machine Learning – Assignment 3

103062528 林玉山

1. 證明 Regularized Linear Regression 的兩種最佳解，可以互相調換：

原式為：

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

先將此式改寫為：

$$\frac{1}{\lambda}\left(\frac{1}{\lambda} X^T X + I_d\right)^{-1} X^\top r$$

根據 Matrix Inversion Lemma

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

將上式之 A,B,C,D 分別代換為：

$$A = I_d, B = X^\top, C = \frac{1}{\lambda} I_n, D = X$$

那我們就可以將原式寫為：

$$\frac{1}{\lambda}(I_d - I_d X^\top (\lambda I_n + X I_d X^\top)^{-1} X I_d) X^\top r$$

將右邊的 $X^\top$ 利用分配律乘入左側

$$= \frac{1}{\lambda}(I_d X^\top - I_d X^\top (\lambda I_n + X I_d X^\top)^{-1} X I_d X^\top) r$$

消除所有可消除的 Identity Matrices

$$= \frac{1}{\lambda}(X^\top - X^\top (\lambda I_n + X X^\top)^{-1} X X^\top) r$$

提出 $X^\top$

$$= \frac{1}{\lambda} X^\top (I_n - (\lambda I_n + X X^\top)^{-1} X X^\top) r$$

因為 $(\lambda I_n + XX^\top)^{-1}\lambda I_n - (\lambda I_n + XX^\top)^{-1}\lambda I_n = 0$

$$= \frac{1}{\lambda} X^\top (I_n - (\lambda I_n + XX^\top)^{-1} XX^\top - (\lambda I_n + XX^\top)^{-1}\lambda I_n + (\lambda I_n + XX^\top)^{-1}\lambda I_n) r$$

提出中間兩式的 $(\lambda I_n + XX^\top)^{-1}$

$$= \frac{1}{\lambda} X^\top (I_n - (\lambda I_n + XX^\top)^{-1}(\lambda I_n + XX^\top) + (\lambda I_n + XX^\top)^{-1}\lambda I_n) r$$

因為 $(\lambda I_n + XX^\top)^{-1}(\lambda I_n + XX^\top) = I_n$

$$= \frac{1}{\lambda} X^\top (I_n - I_n + (\lambda I_n + XX^\top)^{-1}\lambda I_n) r$$

最後整理一下

$$= \frac{1}{\lambda} X^\top ((\lambda I_n + XX^\top)^{-1}\lambda I_n) r$$
$$= X^\top ((\lambda I_n + XX^\top)^{-1}) r$$
$$= X^\top (\lambda I_n + XX^\top)^{-1} r$$

得証。

2. 證明 RKHS 的 inner product 是 well-defined 的

為了要證明，因此必須要達到 symmetry, linearity, positive definiteness 這三項特性

(a) Symmetry

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

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

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

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

(b) Linearity

令 $h = \sum_{l=1}^{p} c^{(l)}k\left(z^{(l)}, \cdot\right)$

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

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

$$= \sum_{l=1}^{p} a\sum_{i=1}^{n}\alpha^{(i)}c^{(l)}k\left(x^{(i)}, z^{(l)}\right) + b\sum_{j=1}^{m}\beta^{(j)}c^{(l)}k\left(y^{(j)}, z^{(l)}\right)$$

$$= a\sum_{l=1}^{p}\sum_{i=1}^{n}\alpha^{(i)}c^{(l)}k\left(y^{(j)}, z^{(l)}\right) + b\sum_{l=1}^{p}\sum_{j=1}^{m}\beta^{(j)}c^{(l)}k\left(y^{(j)}, z^{(l)}\right)$$

$$= a\langle f, h\rangle + b\langle g, h\rangle$$

(c) Positive Definiteness

根據內積的定義：

$$\langle f, f \rangle = \|f\|^2$$

因此，如果 $f(\cdot) = O(\cdot)$

則 $\|f\|^2 = \|0\|^2 = 0$

反之，若 $\langle f, f \rangle = 0$

同樣我們可以得知：

$$\langle f, f \rangle = \|f\|^2 = 0$$

而 $f(\cdot)$ 只可能為 $O(\cdot)$

故得証。

3. 證明這題目中兩個 Hyperplanes 之間的距離為

$$\frac{\mathbf{2}}{\|\boldsymbol{w}\|}$$

已經如果有兩個平行的平面為：

$$w^\top x + b_1 = 0, w^\top x + b_2 = 0$$

則這兩個平面之間的距離為：

$$\frac{|b_1 - b_2|}{\|w\|}$$

若將此題的平面寫成上述形式：

$$w^\top x - b - 1 = 0, w^\top x - b + 1 = 0$$

代入公式後，則可得到兩個平面的距離為：

$$\frac{|-b - 1 + b - 1|}{\|w\|} = \frac{|-2|}{\|w\|} = \frac{2}{\|w\|}$$

故得証。

4. 證明 Semiparametric Representer Theorem

原本的 Representer Theorem 是從 Hypothesis 中挑出一個 g

現在改成 $\tilde{g} = g + b\psi$

在 Representer Theorem 中，我們 g 拆成一個 $g_{\parallel}$ 與 $g_{\perp}$

其中 $g_{\parallel}$ 與 $g_{\perp}$ 分別平行與垂直於 $k\left(x^{(1)},\cdot\right), k\left(x^{(2)},\cdot\right), \ldots, k\left(x^{(n)},\cdot\right)$ 的 span

因此我們也可以把 $\tilde{g}$ 寫為

$$\tilde{g} = g_{\parallel} + g_{\perp} + b\psi$$

又因為 $g_{\perp}$ 垂直，因此與其 $k\left(x^{(1)},\cdot\right), k\left(x^{(2)},\cdot\right), \ldots, k\left(x^{(n)},\cdot\right)$ 的內積皆為 0

所以 $\tilde{g}(x^{(i)})$ 可以寫為

$$\tilde{g}(x^{(i)}) = g_{\parallel}(x^{(i)}) + b\psi$$

假設 Minimizer h 有以下形式：

$$h(x) = \sum_{t=1}^{N} c_t k\left(x^{(t)} \cdot x\right) + h_{\perp}(x) + b\psi$$

由 Representer Theorem 知道，$\tilde{h}(x) = h(x) - h_{\perp}(x)$ 其實比較好

因此最後可以得到最佳的 $\tilde{h}(x)$ 為

$$\tilde{h}(x) = \sum_{t=1}^{N} c_t k\left(x^{(t)} \cdot x\right) + b\psi$$

得証。

5. 送分

(送分)

6. 證明 ROC Curve 下半部的面積剛好是隨機挑選的機率

(不會寫)