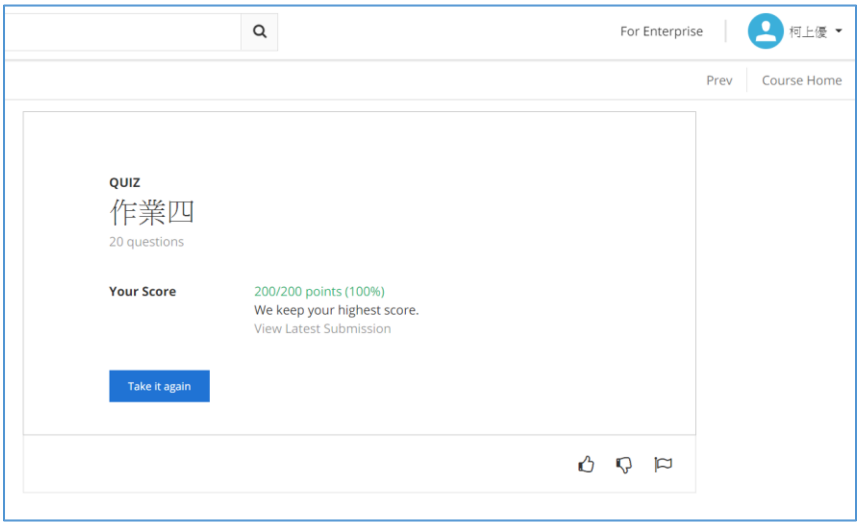
ML homework 4

B05902109 資工二 柯上優　　Submit time: 2018/01/　　:

1.



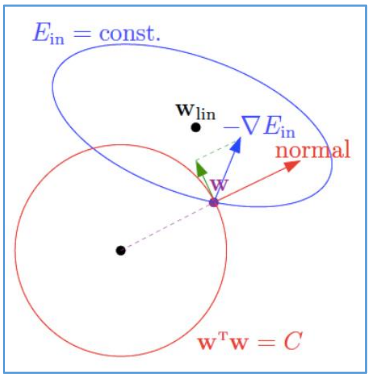
2.

Pf>

$$
\begin{aligned}
w(t+1) &= w(t) - \eta \nabla E_{aug}\big(w(t)\big) \\
&= w(t) - \eta \nabla\left(E_{in}\big(w(t)\big) + \frac{\lambda}{N} w^T w \right) \\
&= w(t) - \eta\left(\nabla E_{in}\big(w(t)\big) + \frac{2\lambda}{N} w(t) \right) \\
&= \left(1 - \eta \frac{2\lambda}{N}\right) w(t) - \eta \nabla E_{in}\big(w(t)\big)
\end{aligned}
$$

3.

由上課投影片可知，$|W_{reg}(\lambda)|$會因為 λ 而變化，若 λ 越小，C 越大，$|W_{reg}(\lambda)|$會增加，但當 C 的範圍包含到 W_{lin}時，$|W_{reg}(\lambda)|$就會停止不繼續和 C 一起成長。因此可知，$|W_{reg}(\lambda)| <= |W_{lin}|$，當 λ >= 0 時，$|W_{reg}(\lambda)|$非遞增，非遞減，非固定

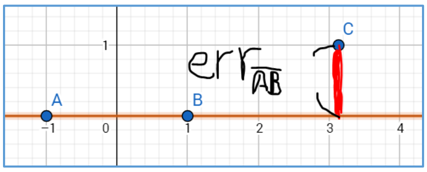


4.

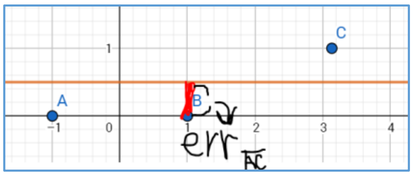
Special thanks : Geogebra

Let A = (-1, 0), B = (1, 0), C = (ρ, 1)

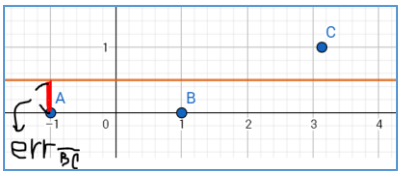
Constant:



Err_{AB} = 1, when we choose C as validation



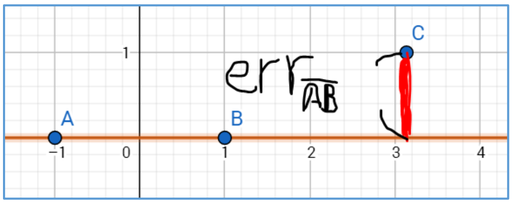
Err_{AC} = $(\frac{1}{2})^2$, when we choose B as validation



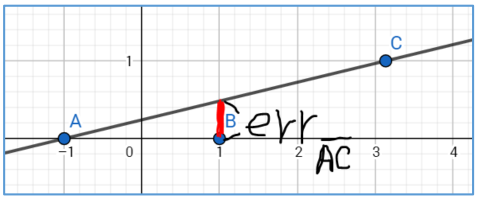
Err_{BC} = $(\frac{1}{2})^2$, when we choose A as validation

So, the $E_{loocv} = \frac{1}{3}(1 + (\frac{1}{2})^2 + (\frac{1}{2})^2)$

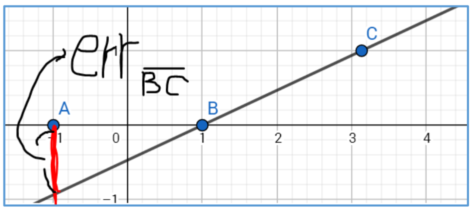
Linear:



Err_{AB} = 1, when we choose C as validation



Err_{AC} = $(\frac{2}{\rho+1})^2$, when we choose B as validation



Err_{BC} = $(\frac{2}{\rho-1})^2$, when we choose A as validation

So, the $E_{loocv} = \frac{1}{3}(1 + (\frac{2}{\rho+1})^2 + (\frac{2}{\rho-1})^2)$

Solve ρ by make two E_{loocv} equal, we can get $\rho = \sqrt{9+4\sqrt{6}}$

5.

題目可視為

$$\min \frac{1}{N+K}\left((w^T X^T X w - 2wX^T y + y^T y) + \left(w^T \tilde{X}^T \tilde{X} w - 2w\tilde{X}^T \tilde{y} + \tilde{y}^T \tilde{y}\right)\right)$$

$$= \min \frac{1}{N+K}\left(w^T w\left(X^T X + \tilde{X}^T \tilde{X}\right) - 2w(X^T y + \tilde{X}^T \tilde{y}) + (y^T y) + (\tilde{y}^T \tilde{y})\right)$$

令

$$\mathrm{A} = \left(X^T X + \tilde{X}^T \tilde{X}\right) \;, b = (X^T y + \tilde{X}^T \tilde{y})$$

By lecture 9, we know optimal linear regression weigh stays in

$$w_{lin} = A^{-1} b^T$$

代入即解。

6.

$$w_{reg} = argmin \;\; \frac{\lambda}{N}\|w\|^2 + \frac{1}{N}\|Xw - y\|^2$$

$$= \mathrm{argmin} \;\; \frac{\lambda}{N} w^T w + E_{in}(w)$$

極值發生於微分等於零

$$\frac{2\lambda}{N}\mathrm{w} + \nabla E_{in}(w) = 0$$

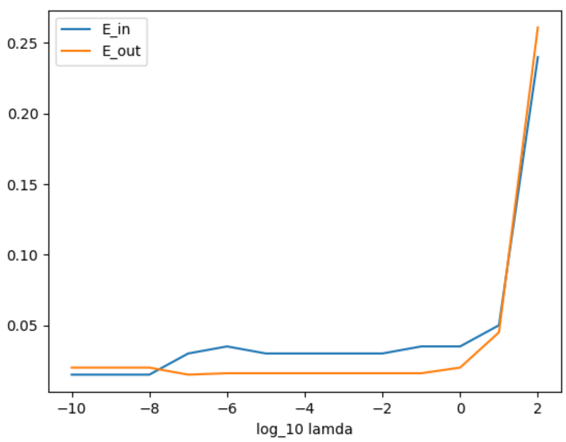
By lecture 14 P.10, 解為

$$(X^T X + \lambda I)^{-1} X^T y$$

和前一題解答比較，解為

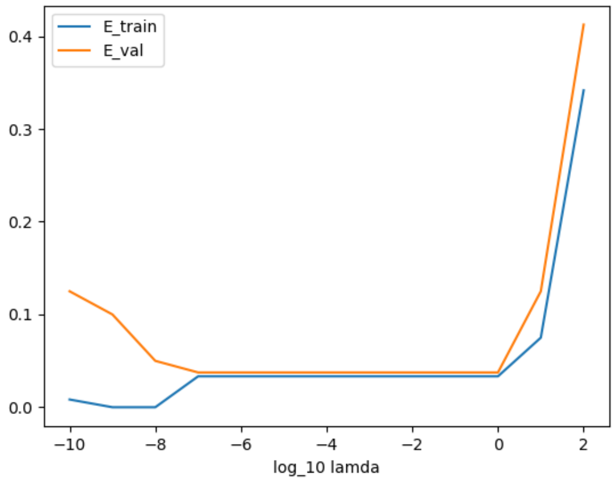
$$\tilde{X} = \sqrt{\lambda}\mathrm{I} \;\; , \tilde{y} = O$$

7.



When log_10 lamda between about -8 and 1, Ein is higher than Eout.
But when it is lower than -8, the constrain C is higher, and Ein and Eout stay small and Ein < Eout.

8.



When log_10 lamda > -7, E_train is smaller than E_validation, showing that the hyposesis perform bad when lamda is too small.

9.

(a)when we choose a instance from positive, the positive in dataset is 1125 and negative is 1126. For majority algorism, it chooses negative set. And for minority algorism, it chooses positive set. So when it comes to calculate E_loocv, majority algorism will consider it negative and E = 1. In contrast, minority algorism will get E = 0. We will choose minority algorism.

(b)

Let y_i 為被選中的

$$\left(\frac{N * \bar{y} - y_i}{N-1} - y_i\right)^2 = \frac{N}{N-1}(\bar{y} - y_i)^2 = \frac{N}{N-1}\sum_{k=1}^{N}(y_k - y_i)^2$$

$$\propto \frac{1}{N-1}\sum_{k=1}^{N}(y_k - y_i)^2 = var$$