ML 2019 r07922003

劉灌慶

1. $$\min_{A,B} F(A,B) = \frac{1}{N}\sum_{n=1}^{N} \ln\left(1+\exp\left(-y_n\left(A\cdot\left(w_{svm}^T \phi(x_n)+b_{svm}\right)+B\right)\right)\right)$$

$$z_n = w_{svm}^T \phi(x_n) + b_{svm}$$

$$p_n = \theta\left(-y_n(Az_n+B)\right)$$

$$\theta(s) = \frac{\exp(s)}{1+\exp(s)} = \frac{1}{\exp^{-1}(s)+1}$$

$$p_n = \frac{1}{\exp^{-1}(-y_n(Az_n+B))+1} = \frac{1}{e^{-(-y_n(Az+B))}+1} = \frac{1}{\exp(y_n(Az+B))+1}$$

$$F(A,B) = \frac{1}{N}\sum_{n=1}^{N}\ln\left(1+\exp\left(-y_n(Az_n+B)\right)\right)$$

$$= \frac{1}{N}\sum_{n=1}^{N}\ln\left(1+\frac{1}{e^{y_n(Az_n+B)}}\right)$$

$$= \frac{1}{N}\sum_{n=1}^{N}\ln\left(\frac{1+e^{y_n(Az_n+B)}}{e^{y_n(Az_n+B)}}\right)$$

$$= \frac{-1}{N}\sum_{n=1}^{N}\ln\left(\frac{e^{y_n(Az_n+B)}}{1+e^{y_n(Az_n+B)}}\right)$$

$$= \frac{-1}{N}\sum_{n=1}^{N}\ln\left(1-\frac{1}{1+e^{y_n(Az_n+B)}}\right)$$

$$= \frac{-1}{N}\sum_{n=1}^{N}\ln(1-p_n)$$

$$\nabla p_n = \nabla\left[e^{y_n(Az+B)}+1\right]^{-1}$$

$$= -\left(e^{y_n(Az+B)}+1\right)^{-2}\nabla\left(e^{y_n(Az+B)}+1\right)$$

$$= -p_n^2 \cdot e^{y_n(Az+B)}\begin{bmatrix} y_n z \\ y_n \end{bmatrix}$$

$$= -p_n^2\left(\frac{1}{p_n}-1\right)\begin{bmatrix} y_n z \\ y_n \end{bmatrix}$$

$$= p_n(p_n-1)\begin{bmatrix} y_n z \\ y_n \end{bmatrix}$$

$$\nabla F(A,B) = \frac{-1}{N}\sum_{n=1}^{N}\frac{1}{1-p_n}\cdot\nabla(1-p_n)$$

$$= \frac{-1}{N}\sum_{n=1}^{N}\frac{1}{1-p_n}(-p_n)(p_n-1)\begin{bmatrix} y_n z \\ y_n \end{bmatrix}$$

$$= \frac{-1}{N}\sum_{n=1}^{N}\begin{bmatrix} p_n y_n z \\ p_n y_n \end{bmatrix} \quad \#$$

2.

$$\frac{\partial F(A,B)}{\partial A} = \frac{-1}{N}\sum_{n=1}^{N} y_n z_n p_n \quad , \quad \frac{\partial P_n}{\partial A} = P_n(P_n-1)y_n z_n$$

$$\frac{\partial F(A,B)}{\partial B} = \frac{-1}{N}\sum_{n=1}^{N} y_n p_n \quad , \quad \frac{\partial P}{\partial B} = P_n(P_n-1)\,y_n$$

$$\frac{\partial F(A,B)}{\partial A^2} = \frac{-1}{N}\sum_{n=1}^{N} y_n z_n \frac{\partial P_n}{\partial A}$$

$$= \frac{-1}{N}\sum_{n=1}^{N} y_n z_n (P_n(P_n-1)y_n z_n)$$

$$= \frac{-1}{N}\sum_{n=1}^{N} y_n^2 z_n^2 P_n(P_n-1)$$

$$= \frac{1}{N}\sum_{n=1}^{N} z_n^2 P_n(1-P_n)$$

$$\frac{\partial F(A,B)}{\partial B^2} = \frac{-1}{N}\sum_{n=1}^{N} y_n \frac{\partial P_n}{\partial B}$$

$$= \frac{-1}{N}\sum_{n=1}^{N} y_n P_n(P_n-1)y_n$$

$$= \frac{1}{N}\sum_{n=1}^{N} y_n^2 P_n(1-P_n)$$

$$= \frac{1}{N}\sum_{n=1}^{N} P_n(1-P_n)$$

$$\frac{\partial F(A,B)}{\partial A \partial B} = \frac{-1}{N}\sum_{n=1}^{N} y_n z_n \frac{\partial P_n}{\partial B}$$

$$= \frac{-1}{N}\sum_{n=1}^{N} y_n z_n P_n(P_n-1)\,y_n$$

$$= \frac{1}{N}\sum_{n=1}^{N} y_n^2 z_n P_n(1-P_n)$$

$$= \frac{1}{N}\sum_{n=1}^{N} z_n P_n(1-P_n)$$

$$\mathcal{H}(F) = \frac{1}{N}\begin{bmatrix} \sum_{n=1}^{N} z_n^2 P_n(1-P_n) & \sum_{n=1}^{N} z_n P_n(1-P_n) \\ \sum_{n=1}^{N} z_n P_n(1-P_n) & \sum_{n=1}^{N} P_n(1-P_n) \end{bmatrix} \quad \#$$

3. $\gamma \to \infty,\ 0 \le \alpha_n \le C,\ \sum_{n=1}^{N} \alpha_n y_n = 0$

Gaussian Kernal : $K(x,x') = \exp(-\gamma \|x-x'\|^2)$

当 $\gamma \to \infty$ , $K(x,x') = [x = x']$

if $x = x'$, $K(x,x') = 1$

if $x \ne x'$, $K(x,x') = 0$

$$\min_{\alpha} \ \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m K(x_n, x_m) - \sum_{n=1}^{N} \alpha_n$$

$$\Rightarrow \min_{\alpha} \ \frac{1}{2}(\alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2) - (\alpha_1 + \alpha_2 + \cdots + \alpha_n)$$

$$\Rightarrow \min_{\alpha} \ \frac{1}{2}(\alpha_1 - 1)^2 + \frac{1}{2}(\alpha_2 - 1)^2 + \cdots + \frac{1}{2}(\alpha_n - 1)^2 + \frac{N}{2}$$

$$\Rightarrow \underbrace{\alpha_1 = \alpha_2 = \cdots = \alpha_n = 1}\ \#$$

4. $(x_1, y_1), (x_2, y_2)$, $y_1 = x_1 - x_1^2$, $y_2 = x_2 - x_2^2$

$$Loss = (w_1 x_1 + w_0 - y_1)^2 + (w_1 x_2 + w_0 - y_2)^2$$

$$\begin{cases} \dfrac{\partial Loss}{\partial w_1} = \cancel{2}(w_1 x_1 + w_0 - y_1)x_1 + \cancel{2}(w_1 x_2 + w_0 - y_2)^2 x_2 = 0 \\ \dfrac{\partial Loss}{\partial w_0} = \cancel{2}(w_1 x_1 + w_0 - y_1) + \cancel{2}(w_1 x_2 + w_0 - y_2) = 0 \end{cases}$$

$$\begin{cases} w_1(x_1^2 + x_2^2) + w_0(x_1 + x_2) - y_1 x_1 - y_2 x_2 = 0 \quad \text{— ①} \\ w_1 \dfrac{(x_1 + x_2)^2}{2} + w_0(x_1 + x_2) - y_1 \dfrac{x_1 + x_2}{2} - y_2 \dfrac{x_1 + x_2}{2} = 0 \quad \text{— ②} \end{cases}$$

① − ② ↓

$$w_1 \frac{(x_1 - x_2)^2}{2} = y_1 x_1 + y_2 x_2 - y_1 \frac{x_1 + x_2}{2} - y_2 \frac{x_1 + x_2}{2} = y_1 \frac{x_1 - x_2}{2} - y_2 \frac{x_1 - x_2}{2}$$

$$w_1 = \frac{y_1 - y_2}{x_1 - x_2} = \frac{x_1 - x_1^2 - x_2 + x_2^2}{x_1 - x_2} = \frac{(x_1 - x_2)(1 - x_1 - x_2)}{x_1 - x_2} = 1 - x_1 - x_2$$

$$w_0 = \frac{y_1 + y_2}{2} - w_1 \frac{x_1 + x_2}{2} = \frac{x_1 x_2}{2} = x_1 x_2$$

$$\bar{g} = \lim_{T \to \infty} \frac{1}{T} \sum g_t = E(1 - x_1 - x_2)x + E(x_1 x_2) = \frac{1}{4}$$

A: $\bar{g}(x) = \frac{1}{4}$

5.

$$\min_w E_{in}^u(w) = \frac{1}{N}\sum_{n=1}^{N} u_n(y_n - w^T x_n)^2$$

$$\Rightarrow \min_w E_{in}^u(w) = \frac{1}{N}\sum_{n=1}^{N}\left(\sqrt{u_n}\, y_n - w^T\sqrt{u_n}\, x_n\right)^2$$

令 $\underline{(\tilde{x}_n, \tilde{y}_n) = (\sqrt{u_n}\, x_n, \sqrt{u_n}\, y_n)}$ #

$$\Rightarrow \min_w E_{in}^u(w) = \frac{1}{N}\sum_{n=1}^{N}\left(\tilde{y}_n - w^T\tilde{x}_n\right)^2$$

6.

$$\frac{u_+^{(2)}}{u_-^{(2)}} = \frac{\varepsilon}{1-\varepsilon} = \frac{22\%}{78\%} = \underline{\frac{11}{39}}$$ #

7. 當 $\theta \geq M$ & $\theta \leq -M$ 時，只有2种可能全為1或−1，跟 $x$ 的維度無關。

當 $-M < \theta < M$，可分成 $M-(-M) = 2M$ 段，每个 $x$ 維為相異，$s = +1, -1$

共 $2 \cdot d \cdot 2M = 4dM$，total $= 4dM + 2$

$d = 2, M = 5$，$2 + 4 \times 2 \times 5 = 42$

$\underline{A: 42}$ #

8.

$$\phi_{ds}(x) = \left(g_1(x), g_2(x), \ldots, g_t(x), \ldots, g_{|g|}(x)\right)$$

$$K_{ds}(x, x') = \left(\phi_{ds}(x)\right)^T\left(\phi_{ds}(x')\right)$$

$$K_{ds} = \text{sign}(x_{i1} - \theta_1)\,\text{sign}(x'_{i1} - \theta_1) + \text{sign}(x_{i2} - \theta_2)\,\text{sign}(x'_{i2} - \theta_2) + \cdots + \text{sign}(x_{i|g|} - \theta_{|g|})\,\text{sign}(x'_{i|g|} - \theta_{|g|})$$

當 $\min(x_{it}, x'_{it}) \leq \theta_t < \max(x_{it}, x'_{it})$ 時，$\text{sign}(x_{it} - \theta_t)\,\text{sign}(x'_{it} - \theta_t) = -1$

其它為 $+1$，$\therefore$ $x_{it}\ x'_{it}$ 決定 1 与 −1 的多寡

$\because s = \pm 1$ $\therefore$ $-1$ 共有 $2\sum_{t=1}^{d}|x_{it} - x'_{it}| = 2\|x - x'\|_1$

$+1$ 共有 $4dM + 2 - 2\|x - x'\|_1$

$$K_{ds}(x, x') = 4dM + 2 - 2\|x - x'\|_1 - 2\|x - x'\|_1$$

$$= \underline{4dM + 2 - 4\|x - x'\|_1}$$ #

# Q9,Q10

λ = [0.05,0.5,5,50,500]

E in = [0.3175, 0.3175, 0.32, 0.315, 0.33]

E out= [0.36, 0.36, 0.36, 0.4, 0.37]

9.
Ans: λ=50 時，有最小的 Ein(g)，Ein(g)=0.315

10.
Ans: λ=0.05,0.5,5 時，有最小的 Eout(g)，Eout(g)=0.36

# Q11,Q12

λ = [0.05,0.5,5,50,500]，選比較好的一次 random

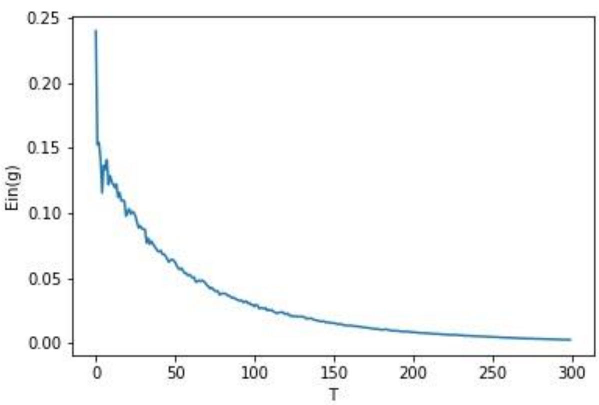
E in = [0.32, 0.32, 0.32, 0.315, 0.3225]

E out= [0.36, 0.36, 0.36, 0.39, 0.37]

11.
Ans: λ=50 時，有最小的 Ein(g)，Ein(g)=0.315
差別在於最 λ=500 時，Ein 縮小了，其他 λ 則是 Ein 上升
12.
Ans: λ=0.05,0.5,5 時，有最小的 Eout(g)，Eout(g)=0.36
在 λ=50，有機會 Eout 比較小

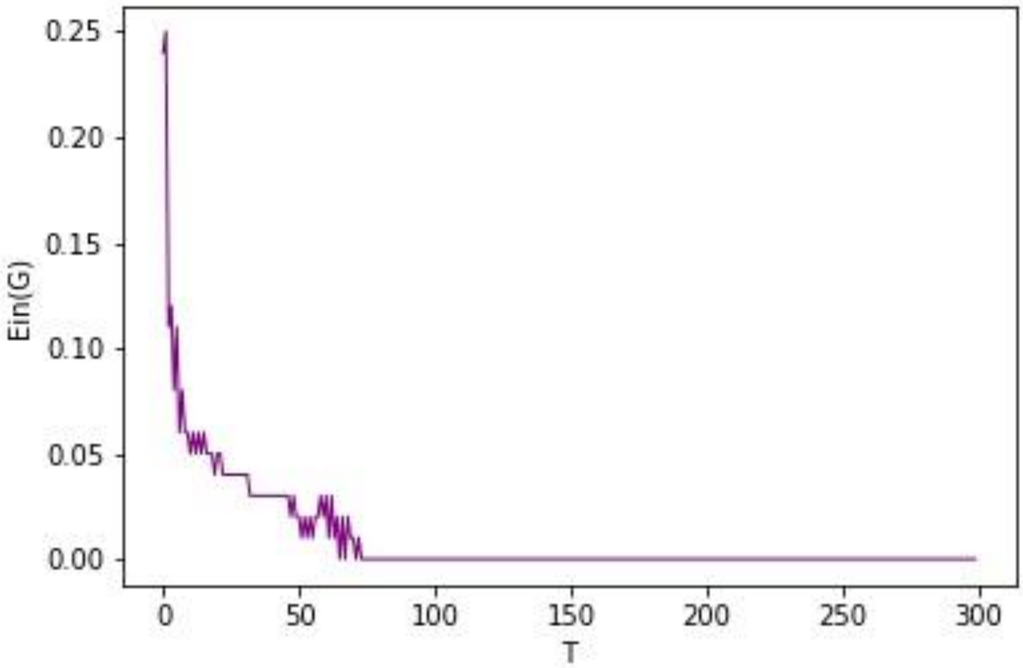
# Q13,Q14,Q15,Q16

13.



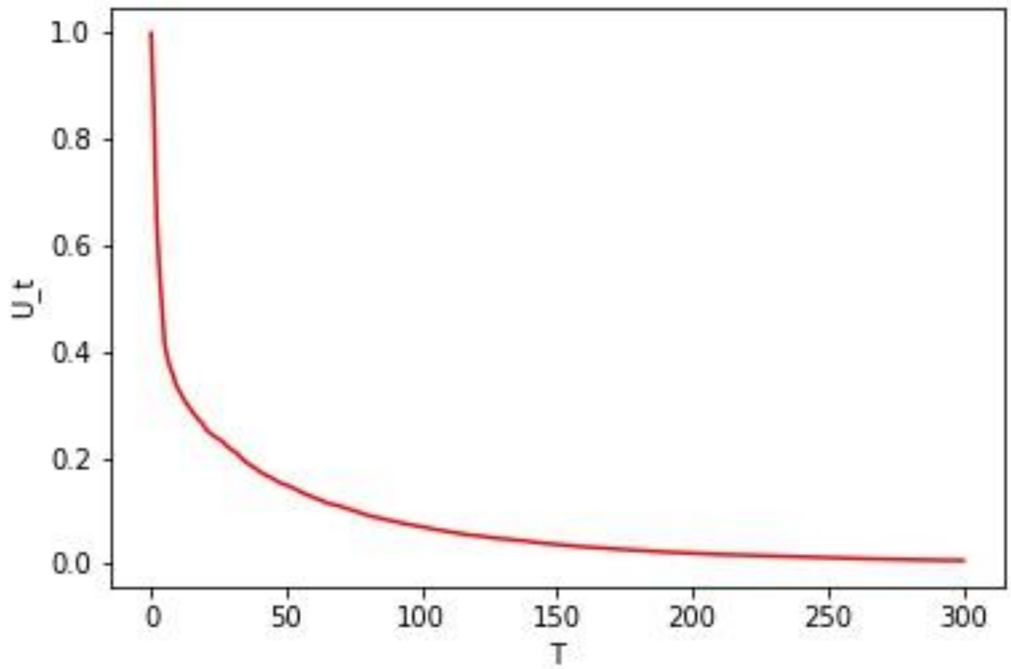

Ein decreasing，如左圖，下降幅度慢慢縮小
Ein(gT)= 0.0026331261839371276

14.



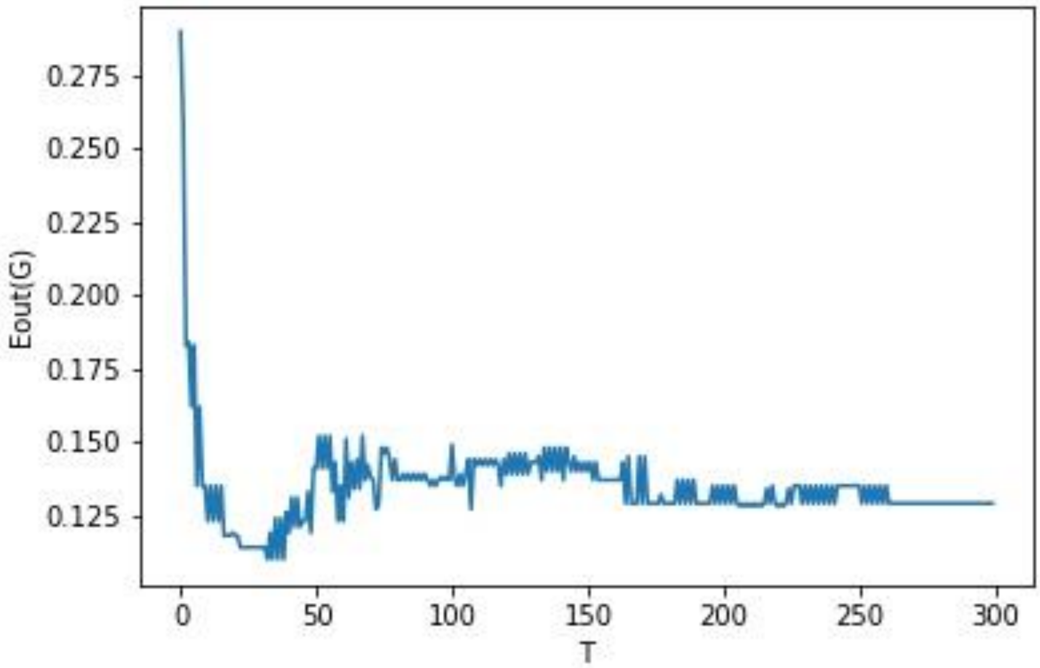

Ein(Gt) 會很快的下降，然後震盪一下再到 0。
到了大概 T=75 就會開始等於 0 了。
Ein(GT)= 0.0。

15.




Ut is decreasing。
Ut 呈現完美的下降圖形，沒有震盪的部分。
U_T= 0.006298022460720289。

16.




Eout(Gt)算呈現 decreasing 的趨勢，但在 T=40 附近時會最小，後面就會回升之後又下降，震盪幅度非常的大，從 T=40 後面開始有點 overfiting 的趨勢。
Eout(T)= 0.129。

$$U_{t+1} = \frac{1}{N}\sum \exp\left(-y_n \sum_{\tau=1}^{t} \alpha_\tau g_\tau(x_n)\right) \quad , \quad \epsilon_t \le \epsilon < \frac{1}{2}$$

17. prove: $U_1 = 1$

$$U_{t+1} = U_t \cdot 2\sqrt{\epsilon_t(1-\epsilon_t)} \le U_t \cdot 2\sqrt{\epsilon(1-\epsilon)}$$

$$U_t = \sum_{n=1}^{N} u_n^{(t)} \quad , \quad u_n^{(t+1)} = \begin{cases} u_n^{(t)} \times \sqrt{\frac{1-\epsilon_t}{\epsilon_t}} & \text{if } y_n \ne g_t(x_n) \\ u_n^{(t)} \Big/ \sqrt{\frac{1-\epsilon_t}{\epsilon_t}} & \text{if } y_n = g_t(x_n) \end{cases} \quad , \quad \epsilon_t = \frac{\sum_{n=1}^{N} u_n^{(t)} [\![ y_n \ne g_t(x_n) ]\!]}{\sum_{n=1}^{N} u_n^{(t)}}$$

$$= u_n^{(t)} \times \left(\sqrt{\frac{1-\epsilon_t}{\epsilon_t}}\right)^{-y_n g_t(x_n)}$$

$$U_{t+1} = \sum_{n=1}^{N} u_n^{(t+1)}$$

$$= \sum_{n=1}^{N} \left( u_n^{(t)} \times \left(\sqrt{\frac{1-\epsilon_t}{\epsilon_t}}\right)^{-y_n g_t(x_n)} \right)$$

$$= \sum_{y_n \ne g_t(x_n)} u_n^{(t)} \times \sqrt{\frac{1-\epsilon_t}{\epsilon_t}} + \sum_{y_n = g_t(x_n)} u_n^{(t)} \times \sqrt{\frac{\epsilon_t}{1-\epsilon_t}}$$

$$= \left(\sum_{n=1}^{N} u_n^{(t)}\right)\left( \sqrt{\frac{1-\epsilon_t}{\epsilon_t}} \times \frac{\sum_{y_n \ne g_t(x_n)} u_n^{(t)}}{\sum_{n=1}^{N} u_n^{(t)}} + \sqrt{\frac{\epsilon_t}{1-\epsilon_t}} \times \frac{\sum_{y_n = g_t(x_n)} u_n^{(t)}}{\sum_{n=1}^{N} u_n^{(t)}} \right)$$

$$= \left(\sum_{n=1}^{N} u_n^{(t)}\right)\left( \sqrt{\frac{1-\epsilon_t}{\epsilon_t}} \times (\epsilon_t) + \sqrt{\frac{\epsilon_t}{1-\epsilon_t}} \times (1-\epsilon_t) \right)$$

$$= U_t \times \left( \sqrt{\epsilon_t(1-\epsilon_t)} + \sqrt{\epsilon_t(1-\epsilon_t)} \right)$$

$$= 2 U_t \sqrt{\epsilon_t(1-\epsilon_t)}$$

$\because \epsilon_t \le \epsilon < \frac{1}{2}$

$\epsilon_t - \epsilon_t^2$

$= -\left(\epsilon_t - \frac{1}{2}\right)^2 + \frac{1}{4}$

$\epsilon_t - \frac{1}{2} \le \epsilon - \frac{1}{2} < 0$

$\left(\epsilon_t - \frac{1}{2}\right)^2 \ge \left(\epsilon - \frac{1}{2}\right)^2 > 0$

$\therefore \epsilon_t(1-\epsilon_t) \le \epsilon(1-\epsilon)$

得證 $\underline{U_{t+1} = U_t \cdot 2\sqrt{\epsilon_t(1-\epsilon_t)} \le U_t \cdot 2\sqrt{\epsilon(1-\epsilon)}}$ #