1.

測驗 • 40 MIN

作業三

向您的目標更進一步
如果您完成此作業，則完成本課程的可能性增加了 **89%**

提交您的作業
截止時間 1月6日 15:59 CST **答題次數** 3/8 hours

再試

收到成績
通過條件 75% 或更高

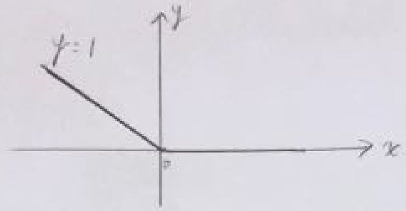
成績
100%

查看反饋

我們會保留您的最高分數

2.

2.



當 $y=1$ 時

若 $w_t^T x<0$,

PLA視此情況 $(x, y=1)$ 是錯的, $\Rightarrow w_{t+1}=w_t+y\cdot x$

此時梯度為 $-x$, $\Rightarrow w_{t+1}=w_t+x=w_t+yx$

若 $w_t^T x\geq 0$,

PLA視此情況 $(x, y=1)$ 是對的 $\Rightarrow w_{t+1}=w_t$

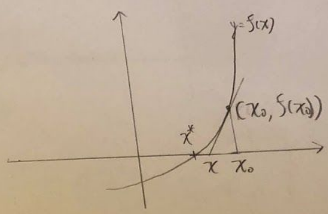
此時梯度為 $\vec{0}$ $\Rightarrow w_{t+1}=w_t+\vec{0}=w_t$

得 $y=1$ 時 PLA與梯度下降相同

同理可證 $y=-1$ 時也成立

3.

3. When Hessian matrix is positive definite, then critical point is local minimum.



$\Rightarrow$ 已知 $f(x) - f(x_0) = f'(x_0)(x - x_0)$

求 $f(x) = 0 \Leftrightarrow -f(x_0) = f'(x_0)(x - x_0)$

同理 $f'(x) = 0 \Leftrightarrow -f'(x_0) = f''(x_0)(x - x_0)$

$\Rightarrow f''(x_0)\,x = f''(x_0)\,x_0 - f'(x_0)$

$$\Rightarrow x = x_0 - \frac{f'(x_0)}{f''(x_0)}$$

又 $f'(x_0) = \nabla E(u,v),\ f'' = \nabla^2 E(u,v)$

$\Rightarrow \underline{x = x_0 - [\nabla^2 E(u,v)]^{-1} \times \nabla E(u,v)}$ #

4.

4、 已知 $h_y(x) = \dfrac{e^{(w_y^T x)}}{\sum_{k=1}^{K} e^{(w_k^T x)}}$

$\Rightarrow Max(P(x_1)h_y(x_1) \cdot P(x_2)h_y(x_2) \cdots P(x_n)h_y(x_n))$,

$\Rightarrow Max(\prod_{i=1}^{n} h(x_i)) \Leftrightarrow Min(-\ln \sum_{i=1}^{n} h_y(x_i))$

$$\Rightarrow -\ln \sum_{i=1}^{n} h_y(x_i) = \sum_{i=1}^{n} \ln\left(\frac{e^{(w_y^T x)}}{\sum_{k=1}^{K} e^{(w_k^T x)}}\right)^{-1}$$

$$= \sum_{i=1}^{n} [\ln(\sum_{k=1}^{K} e^{(w_k^T x)}) - \ln(e^{w_y^T x})]$$

$$= \underline{\sum_{i=1}^{n} [\ln(\sum_{k=1}^{K} e^{(w_k^T x)}) - w_y^T x]}$$ #

5.

5. 假設 X^TX 及 $\tilde{X}^TX$ 皆可逆

已知在使用最小平方誤差搭配線性迴歸時

$$E_{in}(w) = \frac{1}{N+K} \times \left(\sum_{n=1}^{N} (y_n - w^T x_n)^2 + \sum_{k=1}^{K} (\tilde{y}_k - w^T \tilde{x}_k)^2 \right)$$

$$= \frac{1}{N+K} \times \left(\left\| \begin{matrix} x_1^T w - y_1 \\ x_2^T w - y_2 \\ \vdots \\ x_N^T w - y_N \end{matrix} \right\|^2 + \left\| \begin{matrix} \tilde{x}_1^T w - \tilde{y}_1 \\ \tilde{x}_2^T w - \tilde{y}_2 \\ \vdots \\ \tilde{x}_K^T w - \tilde{y}_K \end{matrix} \right\| \right)$$

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

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

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

$$\Rightarrow \nabla E_{in}(w) = \frac{2}{N+K} \left((X^T X + \tilde{X}^T \tilde{X}) w - (X^T y + \tilde{X}^T \tilde{y}) \right)$$

$\Rightarrow w_{lin}$ 存在於 $\nabla E_{in}(w) = 0 = \frac{2}{N+K} \left((X^T X + \tilde{X}^T \tilde{X}) w_{lin} - (X^T y + \tilde{X}^T \tilde{y}) \right)$

$$\Rightarrow (X^T X + \tilde{X}^T \tilde{X}) w_{lin} = X^T y + \tilde{X}^T \tilde{y}$$

$$\Rightarrow \underline{w_{lin} = (X^T X + \tilde{X}^T \tilde{X})^{-1} (X^T y + \tilde{X}^T \tilde{y})} \#$$

6.

6. 承上，已知 $W_{Lin} = (X^TX+\tilde{X}^T\tilde{X})^{-1}(X^Ty+\tilde{X}^T\tilde{y})$

Min $(E_{in}(w) + \frac{\lambda}{N}\|w^2\|)$ $\Leftrightarrow$ 解 $\nabla E_{in}(w) + \frac{2\lambda}{N} w = 0$

$\Rightarrow$ 令 W_{reg} 為上式之解； $\nabla E(W_{reg}) + \frac{2\lambda}{N} w = 0$

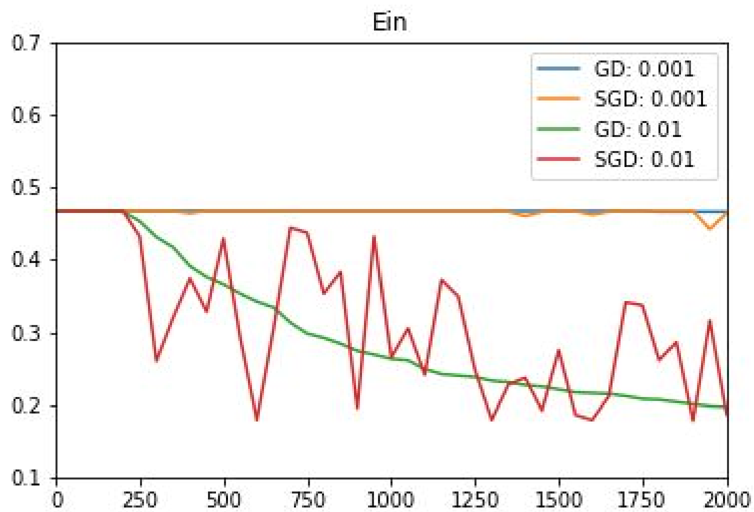
$$\Rightarrow \frac{2}{N}(X^TX W_{reg} - X^Ty) + \frac{2\lambda}{N} W_{reg} = 0$$

$$\Rightarrow (X^TX+I)W_{reg} = X^Ty$$

$$\Rightarrow W_{reg} = (X^TX+I)X^Ty$$

又 $W_{reg} = W_{Lin} \Rightarrow \begin{cases} \tilde{X}^T\tilde{X} = \lambda I \\ \tilde{X}^T\tilde{y} = 0 \end{cases} \Rightarrow \begin{cases} \tilde{X} = \sqrt{\lambda} I \\ \tilde{y} = 0 \end{cases}$

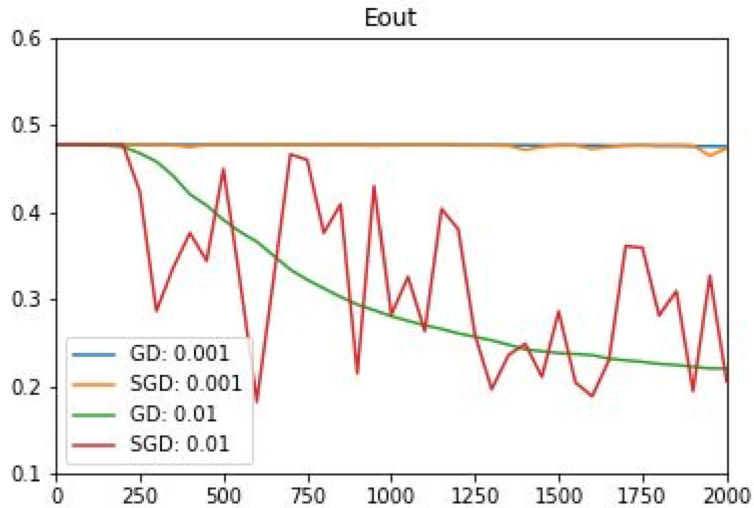
7.
對於 Ein
當 Learning rate 太小時，可能會跨不出 Local Minmum， 而SGD在每個epoch結束後的表現震盪較大， GD則是比較穩定的收斂



8.

對於 Eout 也有跟 Ein 具相同情況

當 Learning rate 太小時，可能會跨不出 Local Minmum， 而SGD在每個epoch結束後的表現震盪較大， GD則是比較穩定的收斂



9.

9(a) 已知 X 經 SVD 得 $X = U\Gamma V^T$

及 $W_{lin} = V\Gamma^{-1}U^T y$

$\Rightarrow X^T X W_{lin}$

$= V\Gamma^T U^T U \Gamma V^T V \Gamma^{-1} U^T y$, $\because V^T V = I, U^T U = I$

$= V\Gamma^T \Gamma \Gamma^{-1} U^T y$

$= V\Gamma^T U^T y$

$= X^T y$ #

(b)