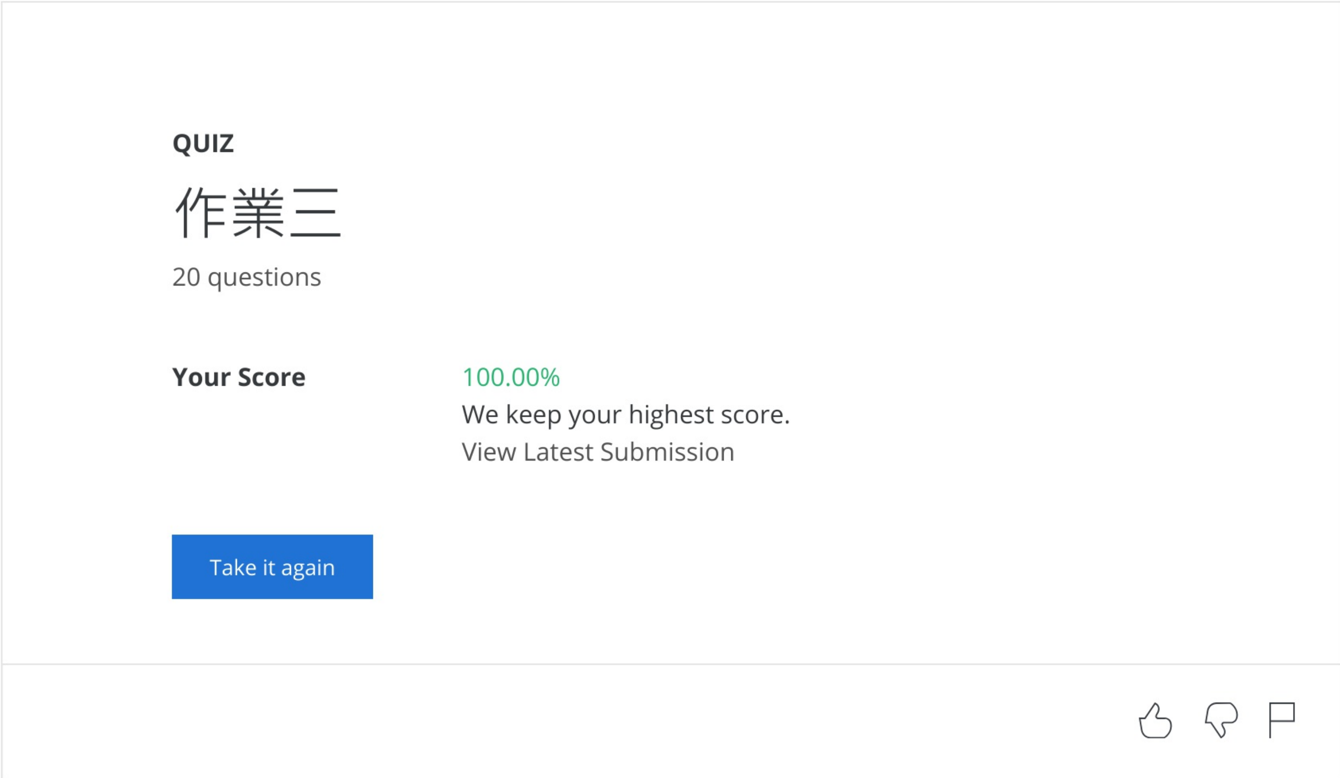
Machine Learning Foundations Homework #3

B04902083 莊翔旭



problem 2

PLA update $\Rightarrow$

$$\begin{aligned} w_{t+1} &\leftarrow w_t + [\![sign(w^T x) \neq y]\!] yx \\ &\leftarrow w_t + 1 \times \begin{cases} 0 & \text{if } sign(w^T x) = y \\ yx = \nabla(-yw^T x) & \text{if } sign(w^T x) \neq y \end{cases} \end{aligned}$$

So SGD $err(w) = max(0, -yw^T x)$ with $\eta = 1$ in PLA.

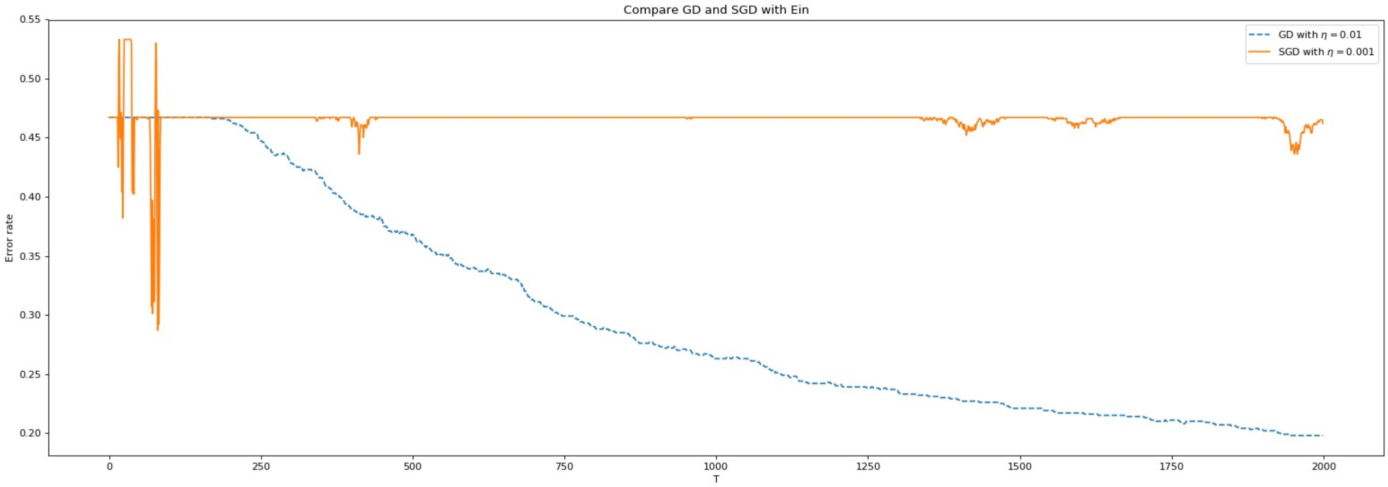
problem 3

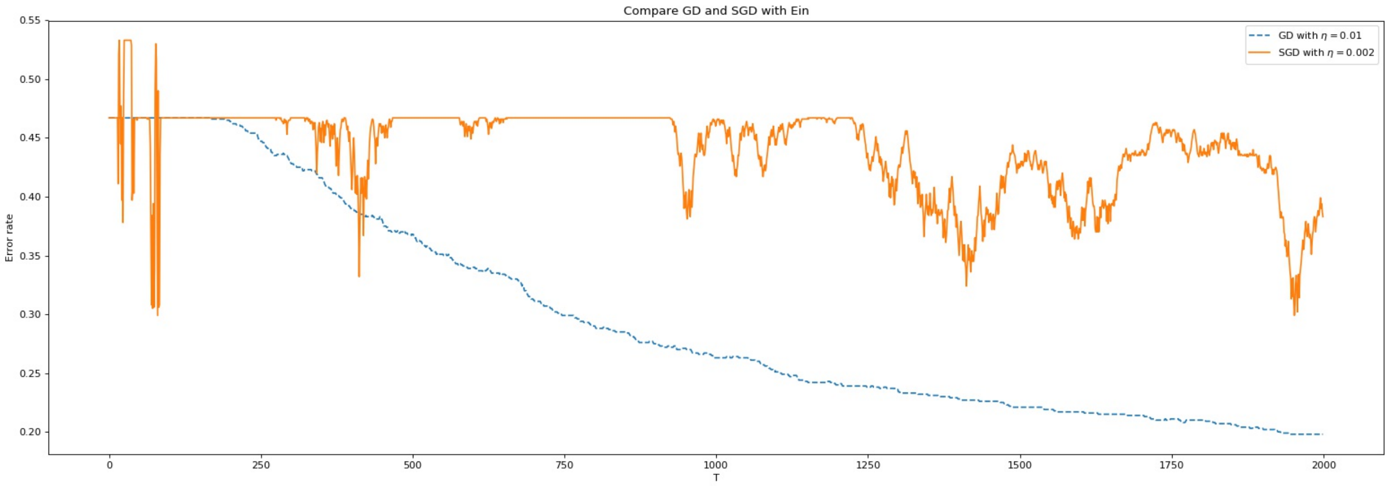
$h_y(x) = \frac{exp(\mathrm{w}_y^T\mathrm{x})}{\sum_{k=1}^K exp(\mathrm{w}_k^T\mathrm{x})}$, and $max\left(\frac{1}{N}\prod_{n=1}^N h_{y_n}(\mathrm{x})\right) \rightarrow min\left(-\frac{1}{N}\sum_{n=1}^N ln(h_{y_n}(\mathrm{x}_n))\right)$

$$
\begin{aligned}
E_{in} &= -\frac{1}{N}\sum_{n=1}^N ln(h_{y_n}(\mathrm{x}_n)) \\
&= -\frac{1}{N}\sum_{n=1}^N ln\left(\frac{exp(\mathrm{w}_{y_n}^T\mathrm{x}_n)}{\sum_{k=1}^K exp(\mathrm{w}_k^T\mathrm{x})}\right) \\
&= -\frac{1}{N}\sum_{n=1}^N ln\left(exp(\mathrm{w}_{y_n}^T\mathrm{x}_n)\right) - ln\left(\sum_{k=1}^K exp(\mathrm{w}_k^T\mathrm{x}_n)\right) \\
&= \frac{1}{N}\sum_{n=1}^N ln\left(\sum_{k=1}^K exp(\mathrm{w}_k^T\mathrm{x}_n)\right) - ln\left(exp(\mathrm{w}_{y_n}^T\mathrm{x}_n)\right) \\
&= \frac{1}{N}\sum_{n=1}^N ln\left(\sum_{k=1}^K exp(\mathrm{w}_k^T\mathrm{x}_n)\right) - \mathrm{w}_{y_n}^T\mathrm{x}_n
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial E_{in}}{\partial \mathrm{w}_i} &= \frac{\partial(\ \frac{1}{N}\sum_{n=1}^N ln\left(\ \sum_{k=1}^K exp(\mathrm{w}_k^T\mathrm{x}_n)\right) - \mathrm{w}_{y_n}^T\mathrm{x}_n\ )}{\partial \mathrm{w}_i} \\
&= \frac{1}{N}\sum_{n=1}^N \frac{\partial(\ ln\left(\ \sum_{k=1}^K exp(\mathrm{w}_k^T\mathrm{x}_n)\right) - \mathrm{w}_{y_n}^T\mathrm{x}_n\ )}{\partial \mathrm{w}_i} \\
&= \frac{1}{N}\sum_{n=1}^N \left(\frac{\partial(\ ln\left(\ \sum_{k=1}^K exp(\mathrm{w}_k^T\mathrm{x}_n)\right)\ )}{\partial \mathrm{w}_i} - \frac{\partial(\ \mathrm{w}_{y_n}^T\mathrm{x}_n\ )}{\partial \mathrm{w}_i}\right) \\
&= \frac{1}{N}\sum_{n=1}^N \left(\frac{\partial(\ ln\left(\ \sum_{k=1}^K exp(\mathrm{w}_k^T\mathrm{x}_n)\right)\ )}{\partial \mathrm{w}_i} - [\![y_n = i]\!] \times \mathrm{x}_n\right) \\
&= \frac{1}{N}\sum_{n=1}^N \left(\frac{1}{\sum_{k=1}^K exp(\mathrm{w}_k^T\mathrm{x}_n)} \times \frac{\partial(\ \sum_{k=1}^K exp(\mathrm{w}_k^T\mathrm{x}_n)\ )}{\partial \mathrm{w}_i} - [\![y_n = i]\!] \times \mathrm{x}_n\right) \\
&= \frac{1}{N}\sum_{n=1}^N \left(\frac{1}{\sum_{k=1}^K exp(\mathrm{w}_k^T\mathrm{x}_n)} \times exp(\mathrm{w}_i^T\mathrm{x}_n) \times \mathrm{x}_n - [\![y_n = i]\!] \times \mathrm{x}_n\right) \\
&= \frac{1}{N}\sum_{n=1}^N \left(\frac{exp(\mathrm{w}_i^T\mathrm{x}_n)}{\sum_{k=1}^K exp(\mathrm{w}_k^T\mathrm{x}_n)} \times \mathrm{x}_n - [\![y_n = i]\!] \times \mathrm{x}_n\right) \\
&= \frac{1}{N}\sum_{n=1}^N \left(h_i(\mathrm{x}) \times \mathrm{x}_n - [\![y_n = i]\!] \times \mathrm{x}_n\right) \\
&= \frac{1}{N}\sum_{n=1}^N \left((h_i(\mathrm{x}) - [\![y_n = i]\!]) \times \mathrm{x}_n\right)_{\sharp}
\end{aligned}
$$

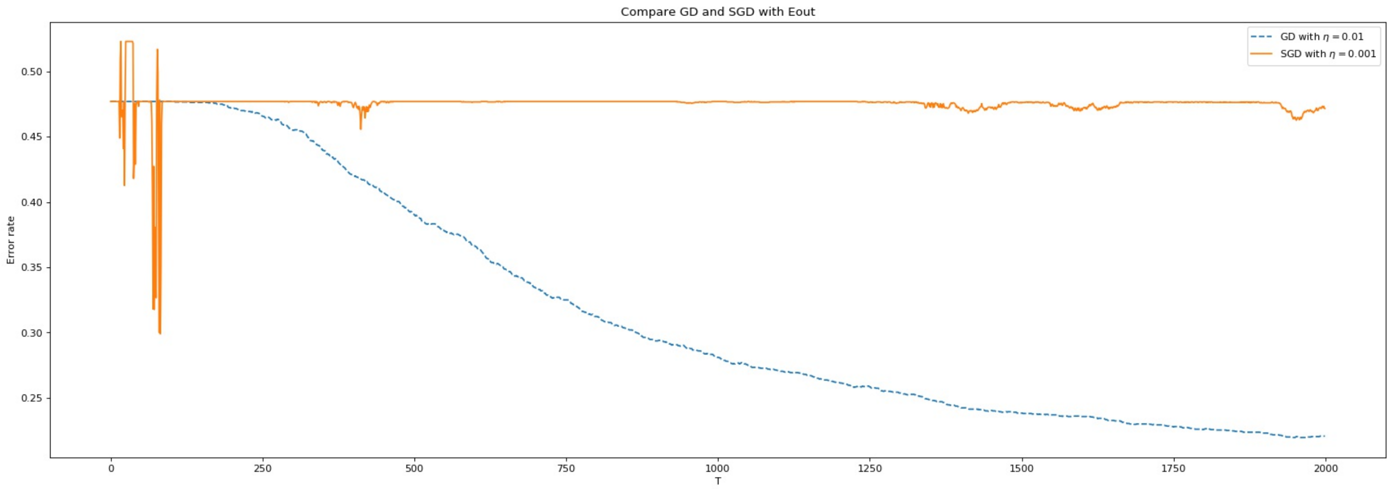
problem 4

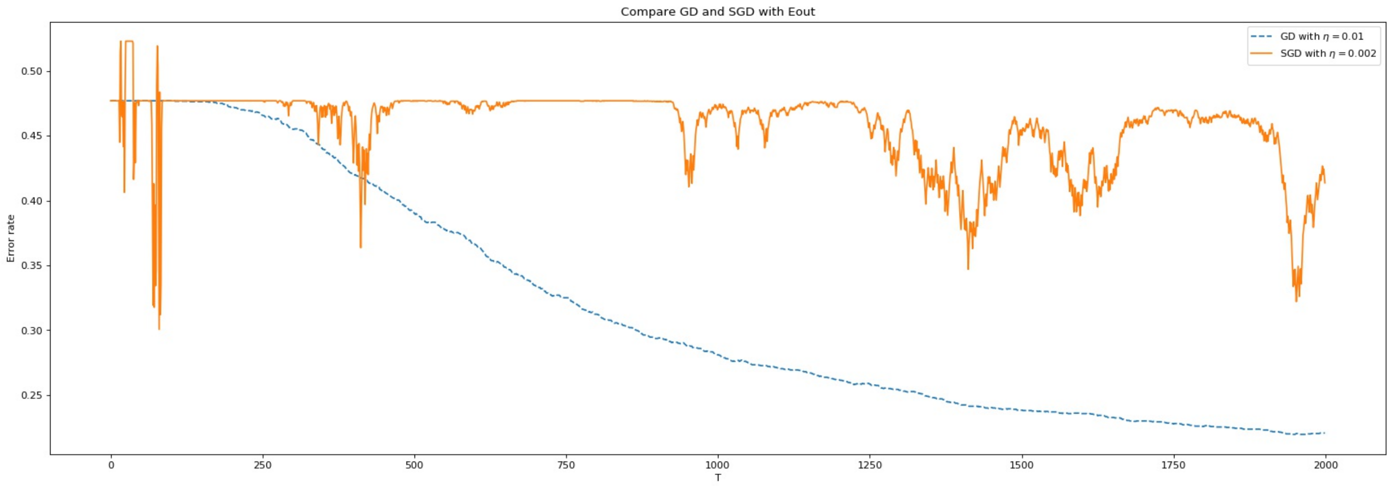




第一張圖可以注意到 GD 的 Error 很會就降下去了， SGD 的 Error 在尾端才有往下降的感覺。為了比較，在畫一張 SGD 在 $\eta = 0.002$ 的圖，可以發現這次 SGD 的 Error 比較早往下降。可以發現在 η 很小的時候會機器學期的確會學的比較慢。

problem 5





可以發現和第四題一樣的結論，且 Eins 和 Eout 不會差太多，表示和預期的一樣有學習到

Bonus