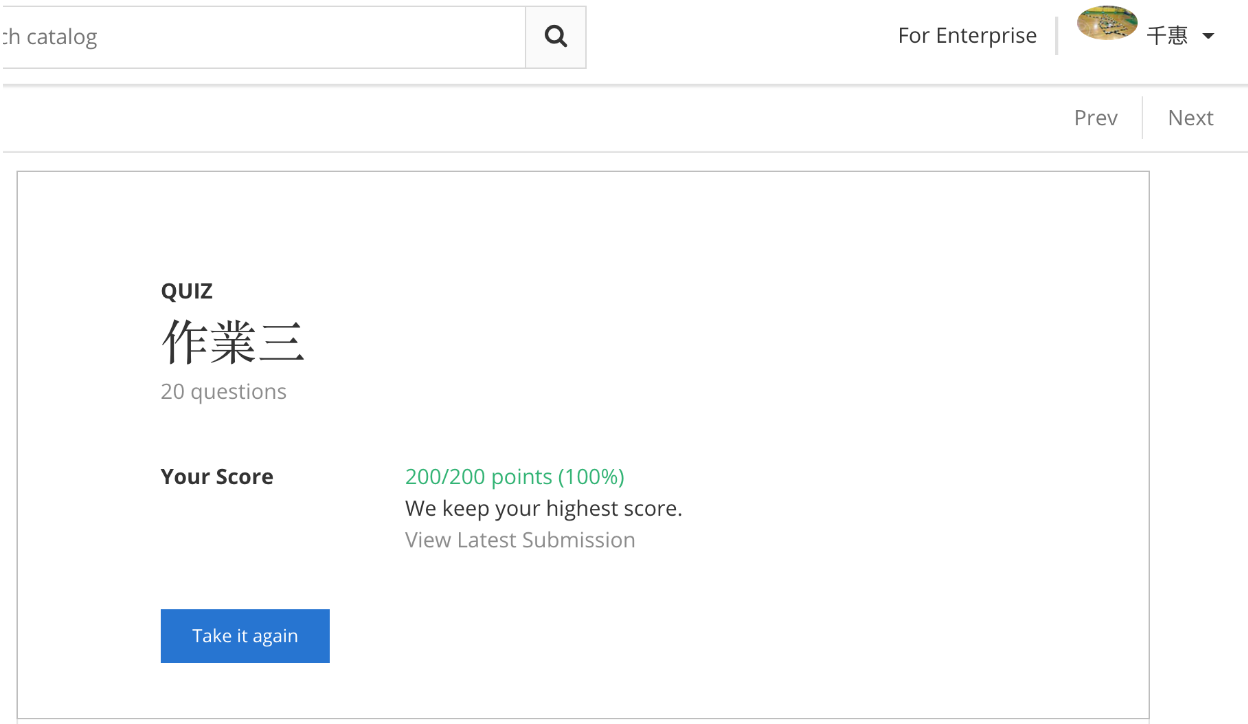
Homework #3

資工三 B04902009 蕭千惠
2nd January, 2018

Problem 1

Score: 200 / 200



Problem 2

$$H = X(X^TX)^{-1}X^T$$

$$\begin{aligned}
H^2 &= X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T \\
&= X(X^TX)^{-1}(X^TX)(X^TX)^{-1}X^T && \text{(By commmutative law.)} \\
&= X(X^TX)^{-1}X^T && (Since(X^TX)^{-1}(X^TX) = I) \\
&= H && \text{(double projection = single one)}
\end{aligned}$$

$$(I-H)^2 = I^2 - 2IH + H^2 = I - 2H + H = I - H \quad \text{(double residual transform = single one)}$$

Problem 3

Prove that SGD with $err(\mathbf{w}, \mathbf{x}, y) = max(0, -y\mathbf{w}^T\mathbf{x})$ results in PLA.

PLA: $\mathbf{w}_{t+1} \leftarrow \mathbf{w}_t + [\![y \neq sign(\mathbf{w}_t^T\mathbf{x})]\!](y\mathbf{x}) = \mathbf{w}_t + (err_{0/1})(y\mathbf{x})$

SGD: $\mathbf{w}_{t+1} \leftarrow \mathbf{w}_t - \nabla err(\mathbf{w}, \mathbf{x}, y)$

When $y = +1$ and $\mathbf{w}_t^T\mathbf{x} > 0$, $\begin{cases} err_{0/1} = 0 \Rightarrow \mathbf{w}_{t+1} \leftarrow \mathbf{w}_t \\ err(\mathbf{w}, \mathbf{x}, y) = 0 \Rightarrow \nabla err(\mathbf{w}, \mathbf{x}, y) = 0 \Rightarrow \mathbf{w}_{t+1} \leftarrow \mathbf{w}_t \end{cases}$

When $y = +1$ and $\mathbf{w}_t^T\mathbf{x} < 0$, $\begin{cases} err_{0/1} = 1 \Rightarrow \mathbf{w}_{t+1} \leftarrow \mathbf{w}_t + \mathbf{x} \\ err(\mathbf{w}, \mathbf{x}, y) = -\mathbf{w}^T\mathbf{x} \Rightarrow \nabla err(\mathbf{w}, \mathbf{x}, y) = -\mathbf{x} \Rightarrow \mathbf{w}_{t+1} \leftarrow \mathbf{w}_t + \mathbf{x} \end{cases}$

When $y = -1$ and $\mathbf{w}_t^T\mathbf{x} > 0$, $\begin{cases} err_{0/1} = 1 \Rightarrow \mathbf{w}_{t+1} \leftarrow \mathbf{w}_t - \mathbf{x} \\ err(\mathbf{w}, \mathbf{x}, y) = \mathbf{w}^T\mathbf{x} \Rightarrow \nabla err(\mathbf{w}, \mathbf{x}, y) = \mathbf{x} \Rightarrow \mathbf{w}_{t+1} \leftarrow \mathbf{w}_t - \mathbf{x} \end{cases}$

When $y = -1$ and $\mathbf{w}_t^T\mathbf{x} < 0$, $\begin{cases} err_{0/1} = 0 \Rightarrow \mathbf{w}_{t+1} \leftarrow \mathbf{w}_t \\ err(\mathbf{w}, \mathbf{x}, y) = 0 \Rightarrow \nabla err(\mathbf{w}, \mathbf{x}, y) = 0 \Rightarrow \mathbf{w}_{t+1} \leftarrow \mathbf{w}_t \end{cases}$

$\Rightarrow$ SGD with $err(\mathbf{w}, \mathbf{x}, y) = max(0, -y\mathbf{w}^T\mathbf{x})$ results in PLA.

Problem 4

To minimize $\hat{E}_2(\Delta u, \Delta v)$, we would ideally like to find Δu and Δv such that $\nabla \hat{E}_2(\Delta u, \Delta v) = 0$.

This is the same as finding $\begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix}$ such that $\nabla E(u + \Delta u, v + \Delta v) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

$E(u + \Delta u, v + \Delta v) \approx \hat{E}_2(\Delta u, \Delta v) = b_{uu}(\Delta u)^2 + b_{vv}(\Delta v)^2 + b_{uv}(\Delta u)(\Delta v) + b_u \Delta u + b_v \Delta v + b$

$$\Rightarrow \nabla E(u + \Delta u, v + \Delta v) \approx \begin{bmatrix} b_u + b_{uu}\Delta u + b_{uv}\Delta v \\ b_v + b_{uv}\Delta u + b_{vv}\Delta v \end{bmatrix} = \begin{bmatrix} b_u \\ b_v \end{bmatrix} + \begin{bmatrix} b_{uu} & b_{uv} \\ b_{uv} & b_{vv} \end{bmatrix} * \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix}$$

$$= \nabla E(u, v) + \nabla^2 E(u, v) \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix}$$

$$0 = \nabla E(u + \Delta u, v + \Delta v) \approx \nabla E(u, v) + \nabla^2 E(u, v) \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix} = -(\nabla^2 E(u, v))^{-1} \nabla E(u, v)$$

Problem 5

$$\text{likelihood(logistic h)} \propto \prod_{n=1}^{N} h_y(\mathbf{x}_n) \propto \ln \prod_{n=1}^{N} h_y(\mathbf{x}_n)$$

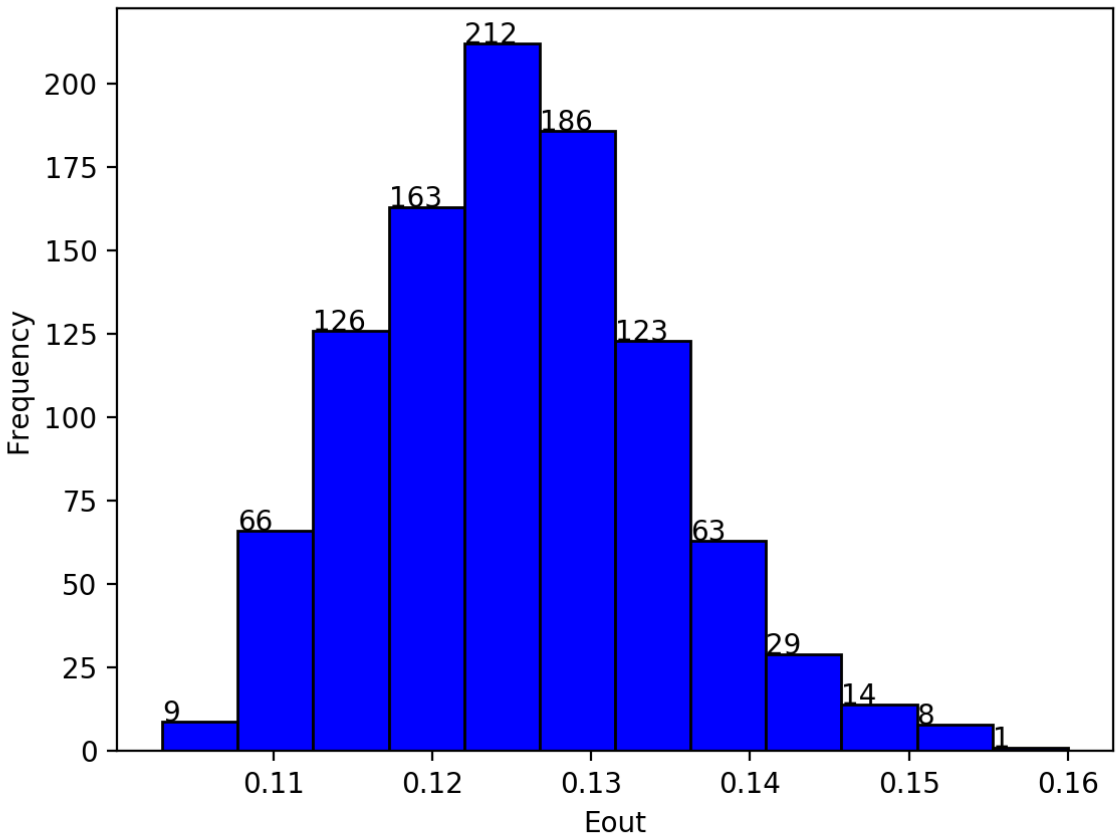
$$\begin{aligned}
\text{max likelihood(logistic h)} &\to \max_h \ \ln \prod_{n=1}^{N} h_y(\mathbf{x}_n) \\
&\to -\min_h \ \frac{1}{N} \ln \prod_{n=1}^{N} h_y(\mathbf{x}_n) \\
&\to -\min \ \frac{1}{N} \ln \prod_{n=1}^{N} \frac{e^{\mathbf{w}_{y_n}^T \mathbf{x}_n}}{\sum\limits_{i=1}^{K} e^{\mathbf{w}_i^T \mathbf{x}_n}} \\
&\to -\min \ \frac{1}{N} \ln \prod_{n=1}^{N} e^{\mathbf{w}_{y_n}^T \mathbf{x}_n} - \ln \prod_{n=1}^{N} (\sum_{i=1}^{K} e^{\mathbf{w}_i^T \mathbf{x}_n}) \\
&\to -\min \ \frac{1}{N} \sum_{n=1}^{N} \mathbf{w}_{y_n}^T \mathbf{x}_n - \sum_{n=1}^{N} \ln \left(\sum_{i=1}^{K} e^{\mathbf{w}_i^T \mathbf{x}_n} \right) \\
&\to \min \ \frac{1}{N} \sum_{n=1}^{N} \left(\ln \left(\sum_{i=1}^{K} e^{\mathbf{w}_i^T \mathbf{x}_n} \right) - \mathbf{w}_{y_n}^T \mathbf{x}_n \right)
\end{aligned}$$

$$E_{in}(\mathbf{w}_1, \mathbf{w}_2, ...\mathbf{w}_K) = \frac{1}{N} \sum_{n=1}^{N} \left(\ln \left(\sum_{i=1}^{K} e^{w_i^T \mathbf{x}_n} \right) - \mathbf{w}_{y_n}^T \mathbf{x}_n \right)$$

Problem 6

$$\begin{aligned}
\frac{\partial E_{in}}{\partial \mathbf{w}_i} &= \frac{\partial \Big[\frac{1}{N} \sum\limits_{n=1}^{N} \left(\ln \left(\sum\limits_{i=1}^{K} e^{\mathbf{w}_i^T \mathbf{x}_n} \right) - \mathbf{w}_{y_n}^T \mathbf{x}_n \right) \Big]}{\partial \mathbf{w}_i} \\
&= \frac{1}{N} \sum_{n=1}^{N} \Big[\Big(\frac{1}{\sum\limits_{i=1}^{K} e^{\mathbf{w}_i^T \mathbf{x}_n}} \cdot e^{\mathbf{w}_i^T \mathbf{x}_n} \cdot \mathbf{x}_n \Big) - [\![y_n = i]\!] \cdot \mathbf{x}_n \Big] \\
&= \frac{1}{N} \sum_{n=1}^{N} \Big((h_i(\mathbf{x}_n) - [\![y_n = i]\!]) \cdot \mathbf{x}_n \Big) \qquad \Big(\text{Since } \frac{1}{\sum\limits_{i=1}^{K} e^{\mathbf{w}_i^T \mathbf{x}_n}} \cdot e^{\mathbf{w}_i^T \mathbf{x}_n} = h_i(\mathbf{x}_n) \Big)
\end{aligned}$$

Problem 7

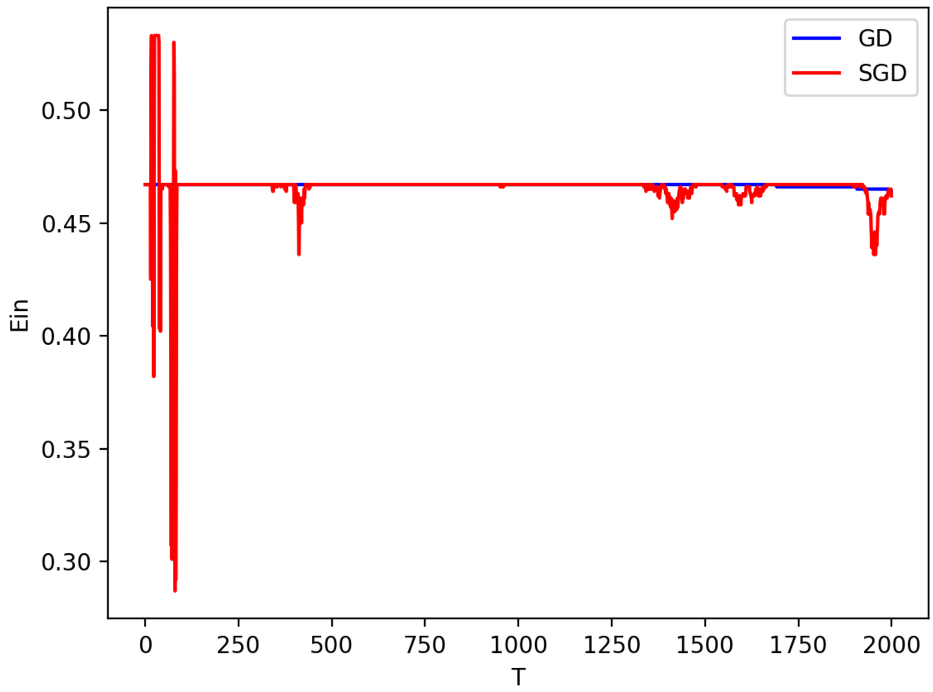


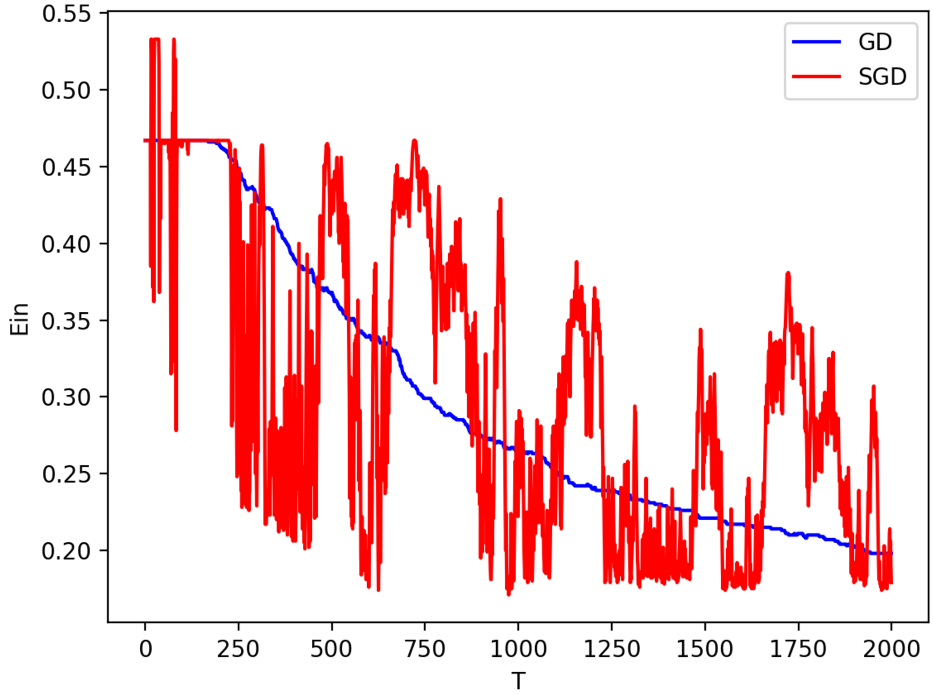
Frequency
Eout
0
25
50
75
100
125
150
175
200
0.11
0.12
0.13
0.14
0.15
0.16
9
66
126
163
212
186
123
63
29
14
8
1

Problem 8

$GD - Ein : 0.467 \rightarrow 0.464$

$SGD - Ein : 0.467 \rightarrow 0.198$

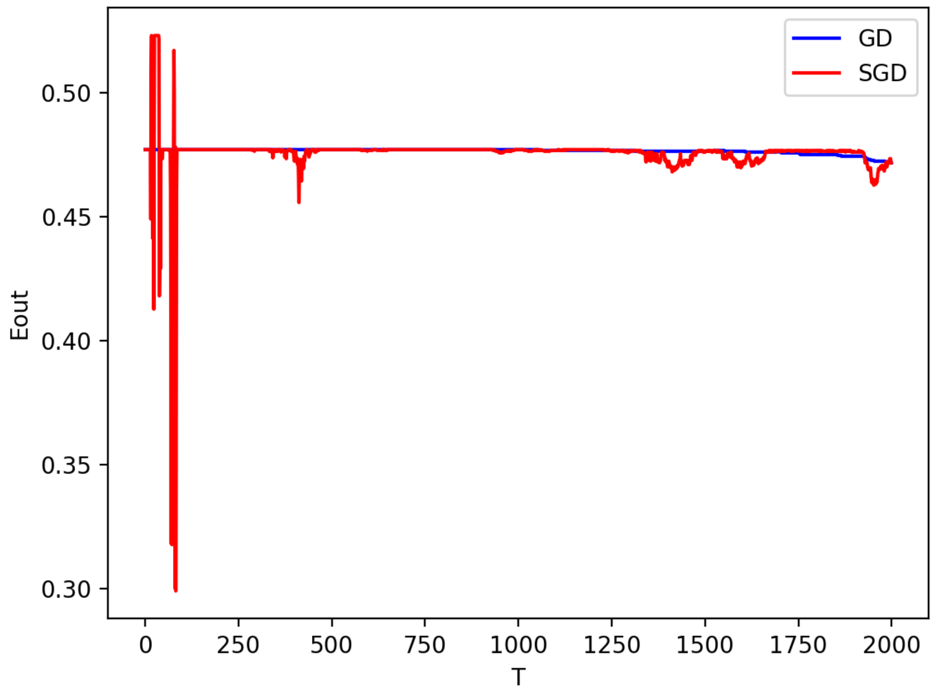


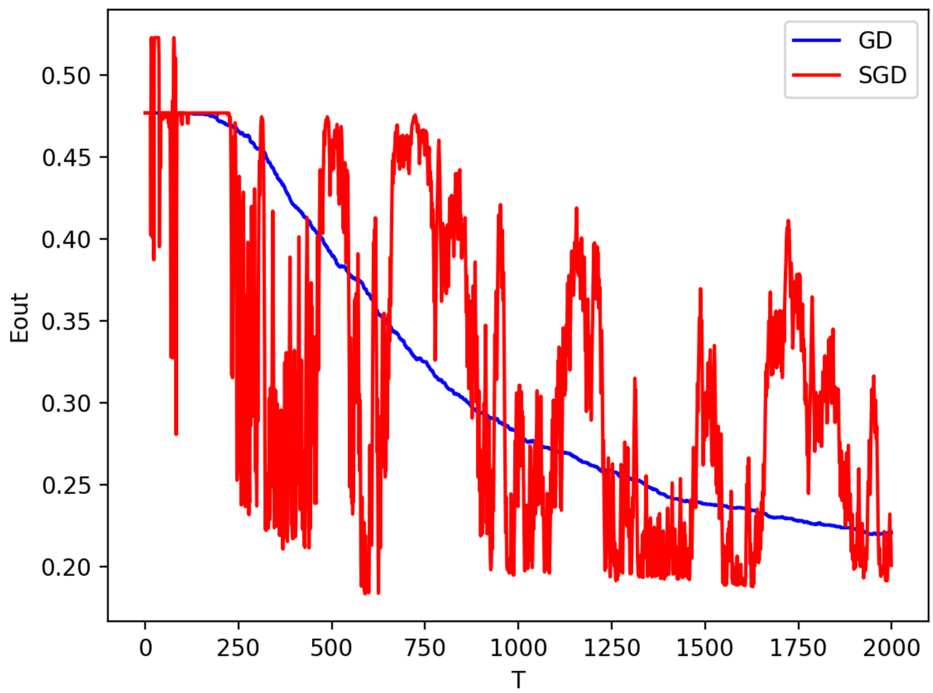


Problem 9

$GD - Eout : 0.477 \rightarrow 0.4716$

$SGD - Eout : 0.477 \rightarrow 0.2206$





Problem 10(Bonus)

(a)

$$
\begin{aligned}
X^T X \mathbf{w}_{lin} &= X^T (U\Gamma V^T)(V\Gamma^{-1}U^T\mathbf{y}) \\
&= X^T U\Gamma(V^T V)\Gamma^{-1}U^T\mathbf{y} && \text{(By commmutative law.)} \\
&= X^T U(\Gamma\Gamma^{-1})U^T\mathbf{y} && \text{(Since } V^T V = I\rho\text{)} \\
&= X^T (UU^T)\mathbf{y} && \text{(Since } \Gamma\Gamma^{-1} = I\rho\text{)} \\
&= X^T\mathbf{y} && \text{(Since } U^T U = I\rho\text{)}
\end{aligned}
$$