ML Foundation: HW3

b04902053 鄭淵仁

December 29, 2017

1

TODO: insert screenshot here.

2

First, we prove $H^2 = H$:

$$\begin{aligned} H^2 &= (X(X^TX)^{-1}X^T)^2 \\ &= (X(X^TX)^{-1}X^T)(X(X^TX)^{-1}X^T) \\ &= X(X^TX)^{-1}[(X^TX)(X^TX)^{-1}]X^T \\ &= X(X^TX)^{-1}X^T \\ &= H \end{aligned}$$

Now we can prove that $(I-H)^2 = (I-H)$:

$$\begin{aligned} (I-H)^2 &= I^2 - 2IH + H^2 \\ &= I - 2H + H \\ &= I - H \end{aligned}$$

3

TODO: 看不懂題目在寫什麼？QQ

4

$$\hat{E}_2(\Delta u, \Delta v) = E(u,v) + \nabla E(u,v) \cdot (\Delta u, \Delta v) + \frac{1}{2}(\Delta u, \Delta v)^T \nabla^2 E(u,v)(\Delta u, \Delta v)$$

Let

$$\begin{cases} 0 = \dfrac{\partial \hat{E}_2(\Delta u, \Delta v)}{\partial \Delta u} = \dfrac{\partial E}{\partial u} + \dfrac{1}{2}\left(2\dfrac{\partial^2 E}{\partial u^2}\Delta u + 2\dfrac{\partial^2 E}{\partial u \partial v}\Delta v\right) \\ \qquad\qquad\qquad\quad = \dfrac{\partial E}{\partial u} + \dfrac{\partial^2 E}{\partial u^2}\Delta u + \dfrac{\partial^2 E}{\partial u \partial v}\Delta v \\ 0 = \dfrac{\partial \hat{E}_2(\Delta u, \Delta v)}{\partial \Delta v} = \dfrac{\partial E}{\partial v} + \dfrac{\partial^2 E}{\partial v^2}\Delta v + \dfrac{\partial^2 E}{\partial v \partial u}\Delta u \end{cases}$$

$$
\begin{cases}
0 = \dfrac{\partial E}{\partial u} + \dfrac{\partial^2 E}{\partial u^2}\Delta u + \dfrac{\partial^2 E}{\partial u \partial v}\Delta v \\
0 = \dfrac{\partial E}{\partial v} + \dfrac{\partial^2 E}{\partial v^2}\Delta v + \dfrac{\partial^2 E}{\partial v \partial u}\Delta u
\end{cases}
$$

We can combine the two equations to only one equation by one vector.

$$
\begin{aligned}
0 &= \nabla E(u,v) + \nabla^2 E(u,v) \cdot (\Delta u, \Delta v) \\
-\nabla^2 E(u,v) \cdot (\Delta u, \Delta v) &= \nabla E(u,v) \\
(\Delta u, \Delta v) &= -\left(\nabla^2 E(u,v)\right)^{-1} \nabla E(u,v)
\end{aligned}
$$

5

$$
\max_h \prod_{n=1}^{N} h_y(x_n) = \max_w \prod_{n=1}^{N} \frac{\exp\left(w_{y_n}^T x_n\right)}{\sum_{k=1}^{K} \exp(w_k^T x_n)}
$$

$$
\begin{aligned}
&\max_w \ln \prod_{n=1}^{N} \frac{\exp\left(w_{y_n}^T x_n\right)}{\sum_{k=1}^{K} \exp(w_k^T x_n)} \\
&= \max_w \sum_{n=1}^{N} \ln \frac{\exp\left(w_{y_n}^T x_n\right)}{\sum_{k=1}^{K} \exp(w_k^T x_n)} \\
&= \max_w \sum_{n=1}^{N} \left(\ln\left(\exp\left(w_{y_n}^T x_n\right)\right) - \ln \sum_{k=1}^{K} \exp\left(w_k^T x_n\right) \right) \\
&= \min_w \sum_{n=1}^{N} \left(\ln \sum_{k=1}^{K} \exp\left(w_k^T x_n\right) - w_{y_n}^T x_n \right)
\end{aligned}
$$

So,

$$
E_{in} = \frac{1}{N} \sum_{n=1}^{N} \left(\ln \sum_{k=1}^{K} \exp\left(w_k^T x_n\right) - w_{y_n}^T x_n \right)
$$

6

$$
\begin{aligned}
&\frac{\partial \left(\sum_{n=1}^{N} \left(\ln \sum_{k=1}^{K} \exp\left(w_k^T x_n\right) \right) \right)}{\partial w_i} \\
&= \sum_{n=1}^{N} \left(\frac{\exp\left(w_i^T x_n\right)}{\sum_{k=1}^{K} \exp(w_k^T x_n)} x_n \right) \\
&= \sum_{n=1}^{N} \left(h_i(x_n) x_n \right)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\frac{\partial E_{in}}{\partial w_i} &= \frac{\partial \left(\frac{1}{N} \sum_{n=1}^{N} \left(\ln \sum_{k=1}^{K} \exp\left(w_k^T x_n\right) - w_{y_n}^T x_n \right) \right)}{\partial w_i} \\
&= \frac{1}{N} \sum_{n=1}^{N} \left((h_i(x_n) x_n) - [[y_n = i]] x_n \right) \\
&= \frac{1}{N} \sum_{n=1}^{N} \left(((h_i(x_n)) - [[y_n = i]]) \, x_n \right)
\end{aligned}
$$

7

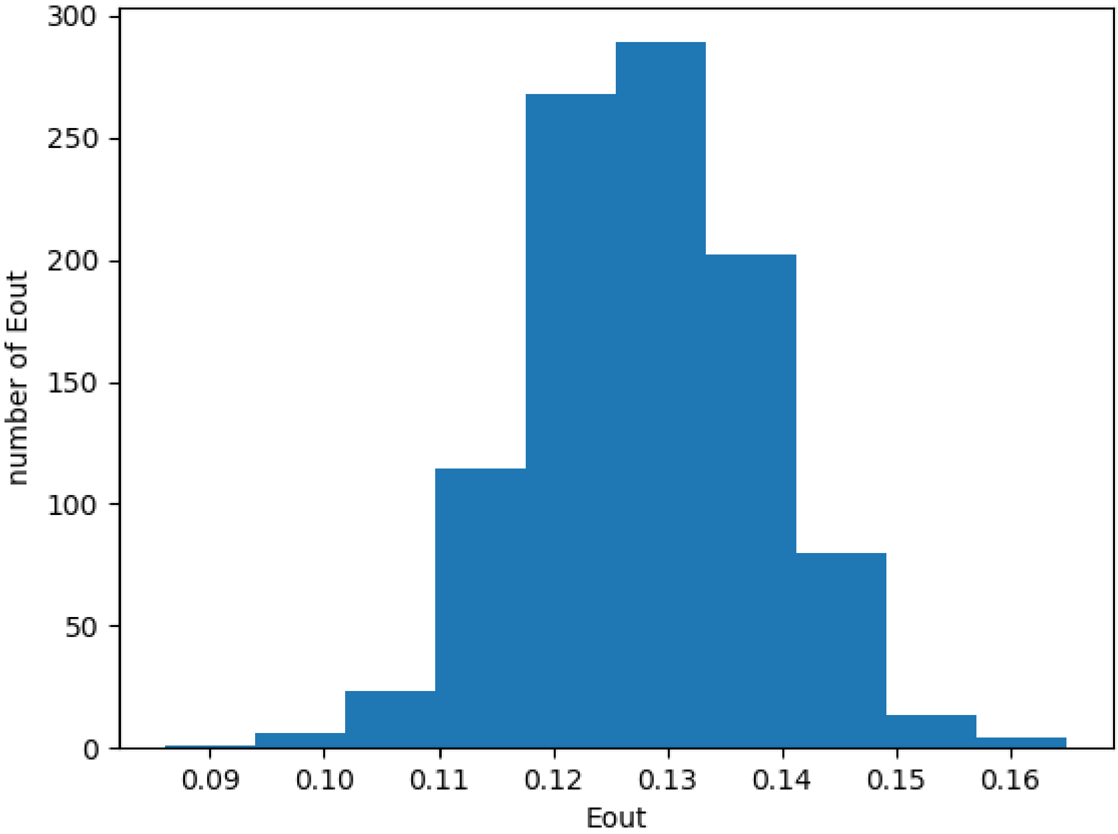

Figure 1: Histogram of E_{out}

Figure 1 shows the histogram of E_{out}.