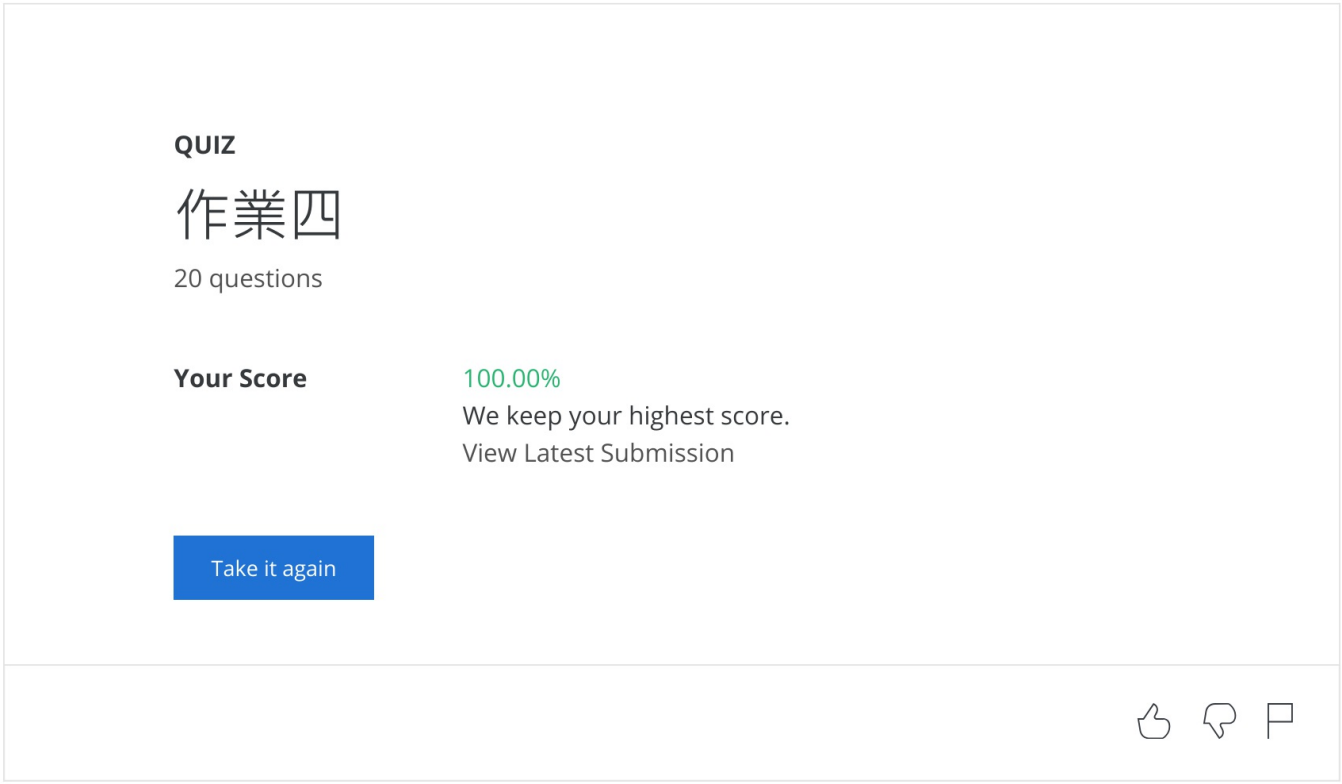
Machine Learning Foundations Homework #4

B04902083 莊翔旭



problem 2

$$\nabla E_{aug}(\mathrm{w}) = \nabla E_{in}(\mathrm{w}) + \frac{2\lambda}{N}\mathrm{w}$$

$$\begin{aligned}
\mathrm{w}_{t+1} &\leftarrow \mathrm{w}_t - \eta \times \nabla E_{aug}(\mathrm{w}_t) \\
&= \mathrm{w}_t - \eta \times (\nabla E_{in}(\mathrm{w}) + \frac{2\eta\lambda}{N}\mathrm{w}_t) \\
&= -\eta \times \nabla E_{in}(\mathrm{w}) + \mathrm{w}_t - \frac{2\eta\lambda}{N}\mathrm{w}_t \\
&= -\eta \times \nabla E_{in}(\mathrm{w}) + (1 - \frac{2\eta\lambda}{N})\mathrm{w}_t
\end{aligned}$$

♯

problem 3

$$
\begin{aligned}
h_1((1,0)) &= \frac{x+1}{\rho+1} && \Rightarrow e_1 = (\frac{2}{\rho+1} - 0)^2 \\
h_1((-1,0)) &= \frac{x-1}{\rho-1} && \Rightarrow e_2 = (\frac{-2}{\rho-1} - 0)^2 \\
h_1((\rho,1)) &= 0 && \Rightarrow e_3 = (0-1)^2
\end{aligned}
$$

$$
\begin{aligned}
E_{loo} &= \frac{1}{3}(e_1 + e_2 + e_3) \\
&= \frac{1}{3}((\frac{2}{\rho+1})^2 + (\frac{-2}{\rho-1})^2 + 1^2)
\end{aligned}
$$

♯

problem 4

```
X = [ x_1, x_2, ..., x_N , xh_1, xh_2, ..., xh_K]
Y = [ y_1, y_2, ..., y_N , yh_1, yh_2, ..., yh_K]

w_i = []
for i in range(T):
    E = randInt(1, N+K)
    w_i.append( (X[E].T * X[E] + λI)^-1 * X[E].T * Y[E] )

w = w_i[-1]
```

$$
\begin{aligned}
P03 \Rightarrow w_t &\leftarrow (1 - \frac{2\eta\lambda}{N})\mathrm{w}_t - \eta \times \nabla E_{in}(\mathrm{w}) \\
P12 \Rightarrow w_t &\leftarrow (x^T x + \lambda I)^{-1} X^T y
\end{aligned}
$$

就像是在投影片說的，SGD 用意在於降低時間複雜度，P03 的 update rule 每次更新都要計算 $\nabla E_{in}(\mathrm{w})$ 為 O(N)，而 P12 的 update rule 是隨機抽一個出來 update，是 O(1)，計算上會比 P03 快。

Bonus