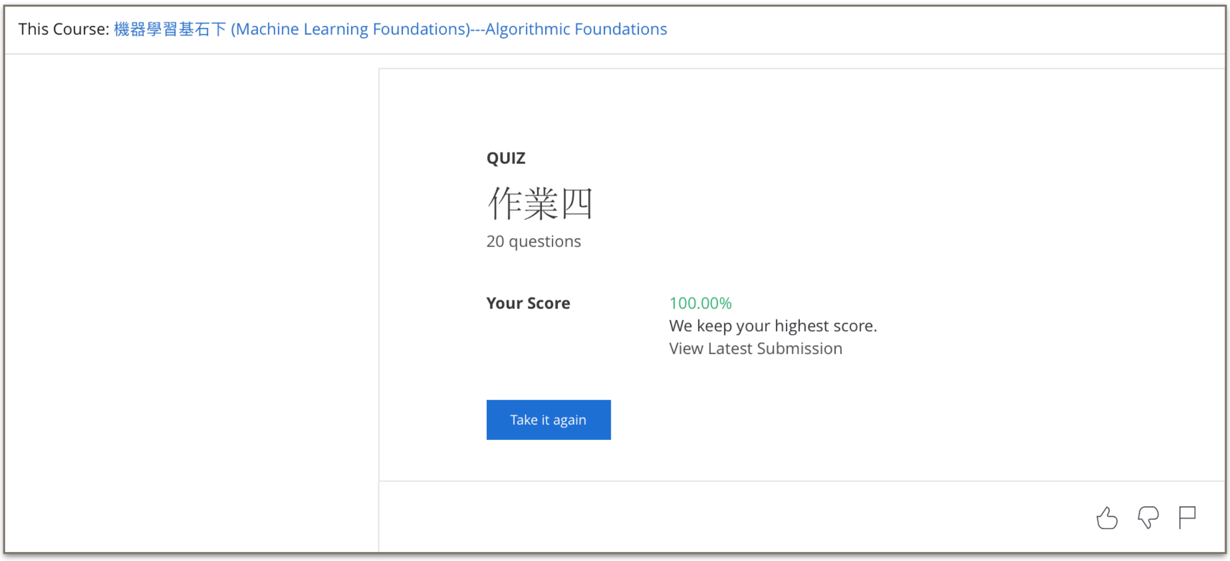
Machine Learning Foundations - *Homework #4*

B05902118　資工三　陳盈如

Problem 1.



Problem 2.

$$
\begin{aligned}
E_{aug}(w) &= E_{in}(w) + \frac{\lambda}{N} w^T w \\
&= E_{in}(w) + \frac{\lambda}{N}(w_0^2 + w_1^2 + \ldots + w_k^2)
\end{aligned}
$$

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

$$
\begin{aligned}
\nabla E_{aug}(w) &= \nabla E_{in}(w) + \frac{2\lambda}{N} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_k \end{bmatrix} \\
&= \nabla E_{in}(w) + \frac{2\lambda}{N} w
\end{aligned}
$$

$$
\begin{aligned}
w_{t+1} &\leftarrow w_t - \eta \cdot \nabla E_{aug}(w_t) \\
&\leftarrow w_t - \eta\left(\nabla E_{in}(w_t) + \frac{2\lambda}{N} w_t \right) \\
&\leftarrow \left(1 - \eta \frac{2\lambda}{N}\right) w_t - \eta \nabla E_{in}(w_t)
\end{aligned}
$$

$$
\underline{w_{t+1} \leftarrow \left(1 - \eta \frac{2\lambda}{N}\right) w_t - \eta \nabla E_{in}(w_t) \#}
$$

Problem 3.

Let two points be (x_1, y_1), (x_2, y_2), and the linear model would be $y = \left(\frac{y_2 - y_1}{x_2 - x_1}\right)x + \left(\frac{x_2y_1 - x_1y_2}{x_2 - x_1}\right)$.

- When two points are $(1,0)$, $(-1,0)$, $y = \left(\frac{0}{2}\right)x + \left(\frac{0}{2}\right)$

  $\rightarrow (\rho,1) : err = (0 + 0 - 1)^2$
- When two points are $(1,0)$, $(\rho,1)$, $y = \left(\frac{1}{\rho - 1}\right)x + \left(\frac{-1}{\rho - 1}\right)$

  $\rightarrow (-1,0) : err = \left(\frac{-1}{\rho - 1} + \frac{-1}{\rho - 1} - 0\right)^2$
- When two points are $(-1,0)$, $(\rho,1)$, $y = \left(\frac{1}{\rho + 1}\right)x + \left(\frac{1}{\rho + 1}\right)$

  $\rightarrow (1,0) : err = \left(\frac{1}{\rho + 1} + \frac{1}{\rho + 1} - 0\right)^2$

$$
\begin{aligned}
E_{loocv}(h_1) &= \frac{1}{3}\left(1 + \left(\frac{-2}{\rho - 1}\right)^2 + \left(\frac{2}{\rho + 1}\right)^2\right) \\
&= \frac{1}{3}\left(1 + \frac{4}{(\rho - 1)^2} + \frac{4}{(\rho + 1)^2}\right)
\end{aligned}
$$

Problem 4.

$$w_{reg} = argmin_w \frac{\lambda}{N}\|w\|^2 + \frac{1}{N}\|Xw - y\|^2$$

$$
\begin{aligned}
\nabla\left(\lambda\|w\|^2 + \|X^Tw - y\|^2\right) &= 2\lambda w + 2\left(X^Tw - y\right)X \\
&= 2\left(\lambda w + X(X^Tw)\right) - 2yX
\end{aligned}
$$

$$
\begin{aligned}
w_{t+1} &\leftarrow w_t - \eta \cdot \nabla\left(\lambda\|w\|^2 + \|X^Tw - y\|^2\right) \\
&\leftarrow w_t - \eta \cdot \left(2\lambda w + 2\left(X^Tw - y\right)X\right) \\
&\leftarrow (1 - 2\eta\lambda)w_t - 2\eta X\left(X^Tw_t\right) + 2\eta yX
\end{aligned}
$$

Findings:

Question 3 uses gradient descent, so it needs to be divided by N; however, Question 4 uses stochastic gradient descent, which takes one vector at a time. Therefore, it is not divided by N.

Problem 5. (Bonus)

$$err(w) = \frac{1}{2\pi}\int_{x=0}^{2\pi} (\sin(\alpha x) - wx)^2 \ dx$$

$$\begin{aligned}\frac{d\ err(w)}{dw} &= \frac{1}{2\pi}\frac{d\int_{x=0}^{2\pi} (\sin(\alpha x) - wx)^2 \ dx}{dw} \\ &= \frac{1}{2\pi}\left(\frac{16\pi^3 w}{3} - \frac{2\big(\sin(2\pi\alpha) - 2\pi\alpha\cos(2\pi\alpha)\big)}{\alpha^2}\right)\end{aligned}$$

$$\frac{16\pi^3 w}{3} - \frac{2\big(\sin(2\pi\alpha) - 2\pi\alpha\cos(2\pi\alpha)\big)}{\alpha^2} = 0$$

$$\frac{16\pi^3 w}{3} = \frac{2\big(\sin(2\pi\alpha) - 2\pi\alpha\cos(2\pi\alpha)\big)}{\alpha^2}$$

$$w = \frac{6\big(\sin(2\pi\alpha) - 2\pi\alpha\cos(2\pi\alpha)\big)}{16\pi^3\alpha^2}$$

$$\int_{x=0}^{2\pi}\left(\sin(\alpha x) - \frac{6\big(\sin(2\pi\alpha) - 2\pi\alpha\cos(2\pi\alpha)\big)}{16\pi^3\alpha^2}x\right) dx = \underline{\frac{4\pi\alpha - 3\sin(2\pi\alpha) + 2\pi\alpha\cos(2\pi\alpha)}{4\pi\alpha^2}\#}$$