Intro to ML 5521 Majid Farhadloo

$$\frac{\partial \text{empirical Loss}}{\partial w_0} = \frac{1}{N}\sum_{t=1}^{N}\left(r^t - (w_1 x^t + w_0)\right)^2$$

(1) $\frac{1}{N}\sum_{t=1}^{N} (r^t)^2 - 2r^t(w_1 x^t + w_0) + (w_1 x^t + w_0)^2$

$\downarrow$

$\left((w_1 x^t)^2 + 2w_1 x^t w_0 + w_0^2\right)$

$\frac{\partial(1)}{\partial w_0}$

$\Rightarrow \frac{1}{N}\sum_{t=1}^{N} -2r^t + 2(w_1 x^t + w_0) \Rightarrow$

(2) $\frac{1}{N}\sum_{t=1}^{N} 2(-r^t + (w_1 x^t + w_0)) = 0$

multiply (2) by $\frac{N}{2} \Rightarrow$

(3) $\sum_{t=1}^{N} w_0 + \sum_{t=1}^{N} w_1 x^t = \sum_{t=1}^{N} r^t$

$\Rightarrow \left(N w_0 + \sum_{t=1}^{N} w_1 x^t = \sum_{t=1}^{N} r^t\right)$

with simillar approach take derivative of (1) with respect to w_1

$\Rightarrow$ (4)

$\Rightarrow \frac{1}{N}\sum_{t=1}^{N} 2(r^t - w_1 x^t - w_0)(-x^t) = 0$

(4) $\times \frac{N}{2}$

$$\Longrightarrow (5) \quad w_0 \sum_{t=1}^{N} x^t + w_1 \sum_{t=1}^{N} (x^t)^2 = \sum_{t=1}^{N} x^t r^t$$

(6) $\rightarrow$ divid 3 by N

$$\Longrightarrow \quad \underline{w_0 + w_1 x^t} = r^t$$

we have two equation of two derivative $\frac{\partial(1)}{w_0}$, $\frac{\partial(1)}{w_1}$

↳ use them to solve equation.

$$\xrightarrow{\times \sum_{t=1}^{N} x^t} \quad N w_0 + w_1 \sum_{t=1}^{N} x^t = \sum_{t=1}^{N} r^t \quad (3)$$

$$\xrightarrow{\times N} \quad w_0 \sum_{t=1}^{N} x^t + w_1 \sum_{t=1}^{N} (x^t)^2 = \sum_{t=1}^{N} x^t r^t \quad (5)$$

$$(7) \quad \left(N w_0 \sum_{t=1}^{N} x^t + w_1 \left(\sum_{t=1}^{N} x^t\right)^2 = \sum_{t=1}^{N} x^t \sum_{t=1}^{N} r^t \right)$$

$$(8) \quad N w_0 \sum_{t=1}^{N} x^t + N w_1 \left(\sum_{t=1}^{N} x^t\right)^2 = N \sum_{t=1}^{N} x^t r^t$$

$(8) - (7) \Longrightarrow$ and then divide by L.H.S

$$w_1 = \frac{N \sum_{t=1}^{N} x^t r^t - \sum_{t=1}^{N} x^t \sum_{t=1}^{N} r^t}{(N-1)\left(\sum_{t=1}^{N} x^t\right)^2}$$

In order to solve for w_0, we need to plugin w_1 and solve equation (6).

The solution for w_1 and w_0 are optimal solution.

(ii) (iii)

$$(1)\quad E(w_2, w_1, w_0 \mid \text{train}) = \frac{1}{N}\sum_{t=1}^{N}\left(r^t - \left(w_2 (x^t)^{2020} + w_1 x^t + w_0\right)\right)^2$$

$$\frac{\partial(1)}{\partial w_0}$$

$$\Rightarrow \frac{1}{N}\sum_{t=1}^{N} (r^t)^2 - 2(r^t)\left(v_2 (x^t)^{2020} + v_1 x^t + v_0\right) + \left(v_2 (x^t)^{2020} + v_1 x^t + v_0\right)^2$$

$$(2)\quad \left((v_2 x^t)^{2020}\right)^2 + (v_1 x^t)^2 + (v_0)^2 + 2 v_2 (x^t)^{2020} v_1 x^t + 2 v_2 (x^t)^{2020} v_0 + 2 v_1 x^t v_0$$

$$\frac{\partial(2)}{\partial v_0} = \frac{1}{N}\sum_{t=1}^{N} 2\left(-r^t + v_2 (x^t)^{2020} + v_1 x^t + v_0\right) = 0$$

$$\Rightarrow \times \frac{N}{2} \Rightarrow N v_0 + v_2 \sum_{t=1}^{N} (x^t)^{2020} + v_1 \sum_{t=1}^{N} x^t = \sum_{t=1}^{N} r^t$$

with simillar approach we can get the derivative w.r.t v_1, v_2.

$$v_1 \Rightarrow v_0 \sum_{t=1}^{N} x^t + v_1 \sum_{t=1}^{N} (x^t)^2 + v_2 \sum_{t=1}^{N} (x^t)^{2021} = \sum_{t=1}^{N} x^t r^t$$

$$v_2 \rightarrow v_0 \sum_{t=1}^{N} (x^t)^2 + v_1 \sum_{t=1}^{N} (x^t)^{2021} + v_2 \sum_{t=1}^{N} (x^t)^{2\times 2020} = \sum_{t=1}^{N} (x^t)^{2020} r^t$$

$$\begin{bmatrix} N & \sum_{t=1}^{N} x^t & \sum_{t=1}^{N} (x^t)^{2020} \\ \sum_{t=1}^{N} x^t & \sum_{t=1}^{N} (x^t)^2 & \sum_{t=1}^{N} (x^t)^{2021} \\ \sum_{t=1}^{N} (x^t)^2 & \sum_{t=1}^{N} (x^t)^{2021} & \sum_{t=1}^{N} (x^t)^{2\times 2020} \end{bmatrix} \begin{bmatrix} v_0 \\ v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{N} r^t \\ \sum_{t=1}^{N} x^t r^t \\ \sum_{t=1}^{N} (x^t)^{2020} r^t \end{bmatrix}$$

in order to solve optimal solutions for v_0, v_1, v_2, it is a good idea that lets matlab or other computer programming language solve it.

(iii))

iii) Yes, his claim is correct, adding more parameter in data model gives us a better fit, and this results in a better model and less error. This could also remind us about overfitting a model.

2.
$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \\ 1 & 4 & 16 & 64 \end{bmatrix}$$

$tr(A) = \underline{76}$

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \\ 1 & 8 & 27 & 64 \end{bmatrix} = A^T \to tr(A^T) = \underline{76}$$

$tr(A^TA) = tr(AA^T) = \boxed{15278}$

$$\begin{bmatrix} 4 & 10 & 30 & 100 \\ 10 & 30 & 100 & 354 \\ 30 & 100 & 354 & 1300 \\ 100 & 354 & 1300 & 4890 \end{bmatrix}$$

ii) as matrix is indicated as 4x4 M, calculting the $|A|$ determines a volume of 4 dimensional "parallelogram" formed by rows of A.

$|A| = \underline{12.000}$

iii) there are multiple ways to derive to solution, one way is to calculate the matrix, and see if the matrix is full rank or not which in this it is as there is no way to derive a column by scale one column + another column and this case rank of matrix is 4 and shows ~~rows~~ columns are linearly independent.

$$\text{rank}(A) = \min(4 \times 4) = \boxed{4}$$

Also, as $|A| \neq 0$ shows that A is not clearly linearly dependent.