Amir Hossein Eyvazkhani

1747696

Ex4

1

a) we can solve that using normal equations:

$$A = X^T \cdot X = \begin{pmatrix} 1.5 & 3 & 4.5 \\ 2 & 2.5 & 3 \end{pmatrix} \begin{pmatrix} 1.5 & 2 \\ 3 & 2.5 \\ 4.5 & 3 \end{pmatrix} = \begin{pmatrix} 31.5 & 24 \\ 24 & 19.25 \end{pmatrix}$$

$$B = X^T y = \begin{pmatrix} 1.5 & 3 & 4.5 \\ 2 & 2.5 & 3 \end{pmatrix} \begin{pmatrix} 10 \\ 15.5 \\ 21 \end{pmatrix} = \begin{pmatrix} 156 \\ 121.75 \end{pmatrix}$$

Solve $Ax = B \longrightarrow x = A^{-1} B = \begin{pmatrix} 2.6 \\ 3 \end{pmatrix} = \beta$

b) first of all, in the part . a), the analytical solution seems easy to find. But it is not the case for all ML problems. Some cases the analetical solution migth be very complex to find. The second reason can be that in such cases, these kinds of solutions migth be computationaly infeasible for current available compute powers. But learning algorithms can be implemented without this issue. Moreover, they can also be designed to be flexible.

c) $L = \| X\beta - y \|_2^2 \longrightarrow \nabla L(\beta) = 2(X^T X \beta - X^T y)$

$\mu = 0.1$, $\beta_0 = (1,1)^T \longrightarrow$ error $= L(\beta) = 108.16$

$\nabla L_1 = \begin{pmatrix} -201 \\ -157 \end{pmatrix}$

$\beta_1 = \beta - \mu \nabla L_1 = \begin{pmatrix} 21.1 \\ 16.7 \end{pmatrix} \longrightarrow$ error $= L(\beta_1) = 8812.7$

$\nabla L_2 = \begin{pmatrix} 1818.9 \\ 1412.25 \end{pmatrix}$

$\beta_2 = \beta_1 - \mu \nabla L_2 = \begin{pmatrix} -160.79 \\ -124.52 \end{pmatrix} \longrightarrow$ error $= 718408.69 = L(\beta_2)$