4) Prediction:- $\hat{y} = sig(X.\theta^T)$

$\therefore X = R^{N \times K}$ & $\theta = R^K$.

where, N = number of inputs

K = # of features.

$\therefore X.\theta^T = O(NK^2)$

& $sig()$ is linear. $\therefore O(NK^2 + N)$.

$\therefore$ Time complexity of Prediction is:-

$$\boxed{O(NK^2)}$$

Space complexity is:- $X = O(NK)$

$\theta = O(K)$.

$\hat{y} = O(N)$.

Since we are performing $X.\theta^T$, $\therefore$ we should concider X in space complexity.

$\therefore$ $$\boxed{SC = O(NK)}$$

$\Rightarrow$ According to some books, inputs are not concidered in ~~Time~~ Space complexity. So,

$$\boxed{SC = O(N + K)}$$

Training.

$$\theta_{K+1} = \theta_K - X^T(y - \hat{y})$$

Time for :-

$\hat{y}$ = $O(NK^2)$

$y - \hat{y}$ = $O(N)$.

$X^T \cdot y$ = $O(KN^2)$

$\theta_K - 0$ = $O(K)$

∴ Time for 1 epoch is :- $O(NK^2 + N + KN^2 + K)$

$$\boxed{TC = O(KN^2)} \quad (\because \text{generally } N > K)$$

Space Complexity :-

We are actually just storing. θ. = $O(K)$ space,

But according to some books, since are using X & y for calculation & hence we are temporarily storing it.

$$\boxed{\therefore SC = O(NK + N + K)}$$