Linear Regression:

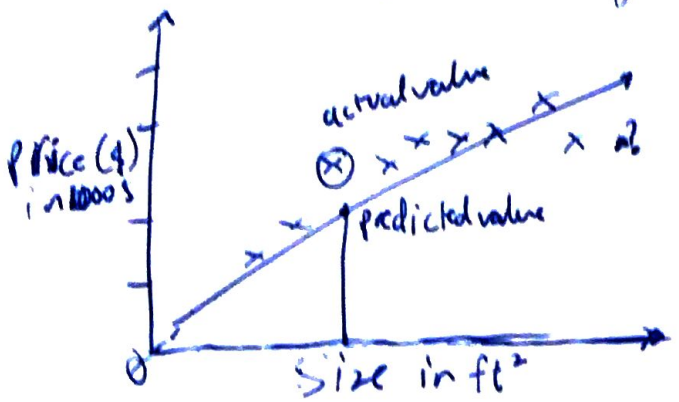
→ In Regression, $f(x)$ & $\therefore g(x)$ returns a continuous value.

→ let $g(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots \theta_D x_D$

→ What are the params $\theta = [\theta_0, \theta_1, \theta_2 \dots \theta_D]^T$ of model?

$(D+1)$ coefficients

such that $g(x)$ works well in predicting target values

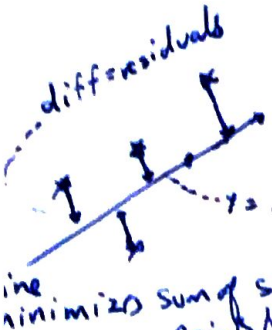


error $= |y_i - g(x_i)|$ in a value

accumulated error $= \sum_{i=1}^{N} (|Y_i - g(x_i)|)^2$

we use ~~differentiable~~ squared value

$$\theta = [\theta_0, \theta_1]^T \text{ such that } g(x_i) \approx Y_i \ \forall (x_i, Y_i)$$



$$J = \sum (Y_i - \hat{Y}_i)^2$$

we want to minimize J

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2 \quad (\hat{Y}_i = g(x_i, \theta))$$

bias = the 1st feature that we put to make a whole matrix

$(Y - X\theta)$ ← $(D+1) \times 1$ col vector

$N \times 1$ ↑ $N \times (D+1)$

$$J = (Y - X\theta)^T (Y - X\theta)$$

Minimize this?
→ take derivative wrt θ
→ set equal to 0 & solve for θ

$$\frac{\partial J}{\partial \theta} = 0$$