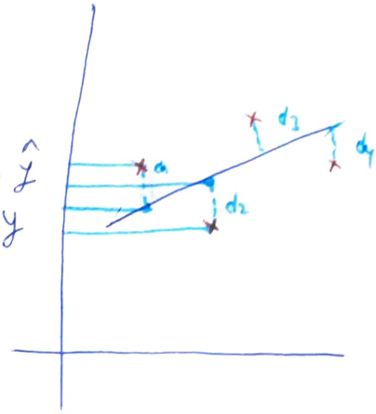
Pts. far away have higher contribution to error.

Derivative of modulus doesn't exist. → ∴ L_1 norm is not used.
L_2 is used

$$d_i^2 = (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2$$

x remains same for all pts., but y change.

Pred - $\hat{y}$
GT : y



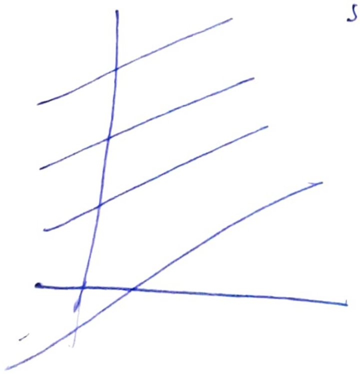
$$\therefore d_i^2 = (y_i - \hat{y}_i)^2$$

$$\therefore J = \sum_{i=1}^{n} (\underbrace{y_i}_{\text{from data}} - \underbrace{\hat{y}_i}_{wx_i + b \text{ (pred)}})^2$$

$$J(w, b) = \sum_{i=1}^{n} (y_i - wx_i - b)^2$$

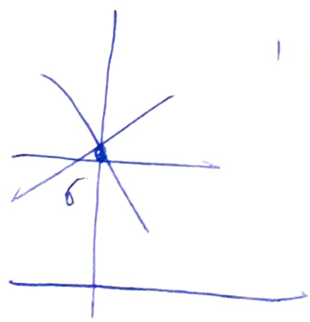
① If w is constant,

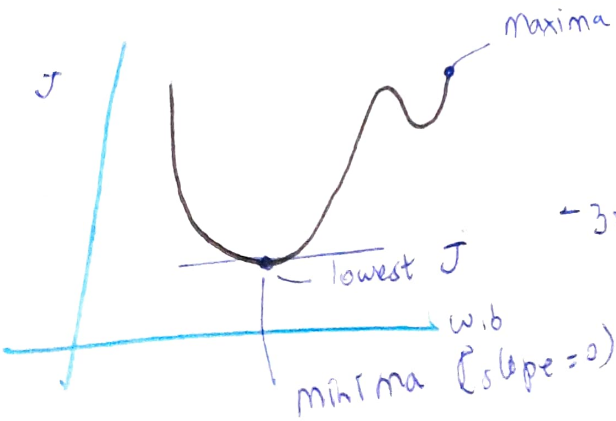
slope is constant, but diff intercepts



② If b is constant.

1 intercept, diff. slopes





Goal- min. J

← 3D

If slope $= 0$, minima is found.

$$\frac{\partial E}{\partial w} = 0 \qquad \frac{\partial E}{\partial b} = 0$$

$$w = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$b = \bar{y} - w\bar{x}$$

$\bar{x}, \bar{y} \rightarrow$ mean.