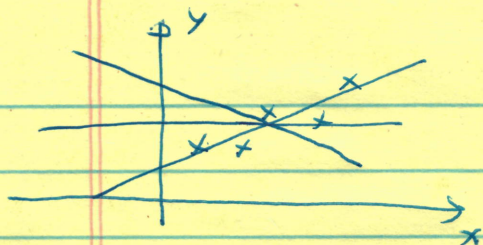


Use x to predict y

$(x_1, y_1), \dots, (x_n, y_n)$



Hypothesis: linear regression

$$y = ax + b$$

$$\hat{y}_i = ax_i + b$$

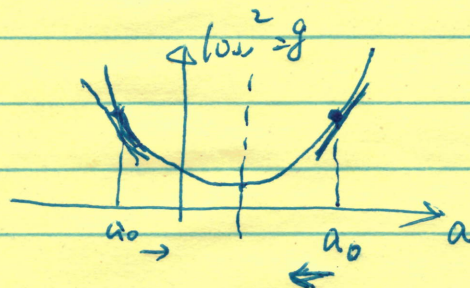
a, b
weight
parameters

$$\text{loss} = \text{rmse} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

$$J = \text{loss}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - ax_i - b)^2$$

$$a_1 = a_0 - \frac{\partial J}{\partial a} \bigg|_{a_0} * \text{lr}$$

gradient descent



$$b_1 = b_0 - \frac{\partial J}{\partial b} \bigg|_{b_0} * \text{lr}$$

lr: learning rate

Three things

- weights (parameters) = a, b
- loss function / objective function / cost function
- optimizer: gradient descent