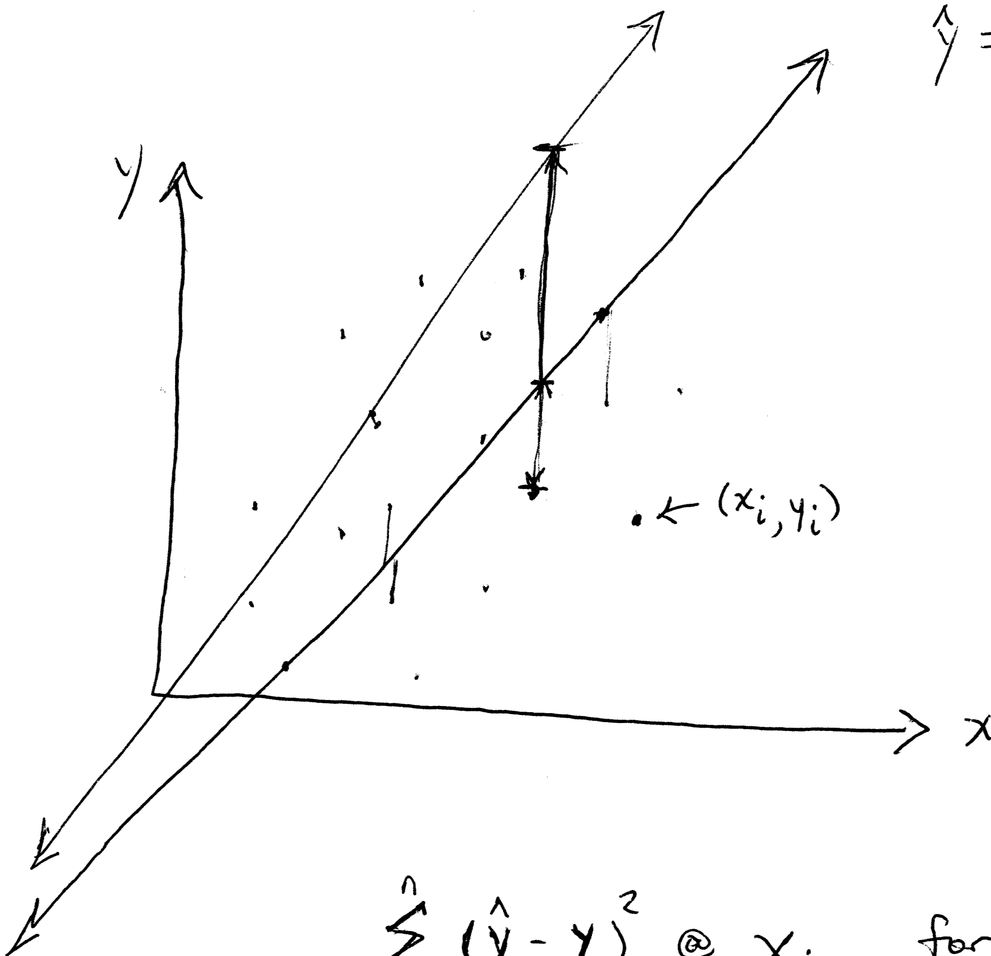
# Linear Regression

Eugenia Duncan

$$\hat{y} = m\hat{x} + b$$




$$\sum_{i=1}^{n} (\hat{y} - y_i)^2 \ @ \ x_i \quad \text{for } m, b$$

Goal is to Minimize Error (Find line with minimum error)

Use Gradient Descent

min of 2 pts. to make a line: $\left\{ \begin{array}{l} (0,-2)^{\text{y-int}} \\ (3, 4) \end{array} \right.$

also $(1, 0)^{\text{x-int}}$ on line

$y = mx + b$

$4 = m(3) - 2$

$6 = 3m$

$m = 2$

$y = 2x - 2$

Exact Line

$\left. \begin{array}{l} (1, 2) \\ (3, 1) \\ (1, -1) \end{array} \right\}$ Additional points

with additional points, what is new eqn of the line? (new m & b)

Find partial derivatives of Sum of Squares Error, SSE, with respect to both m & b, using the Chain Rule for derivatives...

Let SSE $= Q$

$$SSE = \sum_{i=1}^{n} (y_i - (mx_i + b))^2 = Q$$

---

* $$\frac{\partial Q}{\partial b} = \frac{\partial Q}{\partial w} \cdot \frac{\partial w}{\partial b}$$

$w = y_i - (mx_i + b) = y_i - mx_i - b$

$Q = w^2$ $\qquad \frac{\partial Q}{\partial w} = 2w$

$\frac{\partial w}{\partial b} = -1$ $\qquad \frac{\partial Q}{\partial b} = \sum 2(y_i - mx_i - b) \cdot (-1)$

b is the variable
everything else = constant

$$= -2 \sum (y_i - (mx_i + b))$$

---

* $$\frac{\partial Q}{\partial m} = \frac{\partial Q}{\partial w} \cdot \frac{\partial w}{\partial m}$$

$w = y_i - mx_i - b$

$\frac{\partial w}{\partial m} = -x_i$ $\qquad \frac{\partial Q}{\partial m} = \sum 2(y_i - mx_i - b) \cdot (-x_i)$

m is the variable
everything else = constant

$$= -2 \sum x_i (y_i - (mx_i + b))$$

Set the partial derivatives to zero...

$$\left( SSE = \sum_{i=1}^{n} (y_i - (mx_i + b))^2 = Q \right)$$

$\frac{\partial Q}{\partial b}$

$$-2 \sum (y_i - mx_i - b) = 0$$

$$\sum (y_i - mx_i - b) = 0$$

$$\sum y_i - \sum mx_i - \sum b = 0$$

$$\sum y_i - \sum mx_i - nb = 0$$

$$\sum y_i - m \sum x_i - nb = 0$$

$$\sum y_i = nb + m \sum x_i \checkmark$$

$\frac{\partial Q}{\partial m}$

$$\sum x_i y_i - \sum m x_i^2 - \sum b x_i = 0$$

$$\sum x_i y_i - m \sum x_i^2 - b \sum x_i = 0$$

$$\sum x_i y_i = b \sum x_i + m \sum x_i^2 \checkmark$$

Solve for the slope, m, using substitution...

$$m = \frac{\Sigma y_i - nb}{\Sigma x_i}$$

$$m = \frac{\Sigma y_i - n\left(\frac{\Sigma x_i y_i - m \Sigma x_i^2}{\Sigma x_i}\right)}{\Sigma x_i}$$

$$m\Sigma x_i = \Sigma y_i - \frac{n \Sigma x_i y_i}{\Sigma x_i} + \frac{n m \Sigma x_i^2}{\Sigma x_i}$$

$$\frac{m\Sigma x_i \Sigma x_i}{\Sigma x_i} = \frac{\Sigma x_i \Sigma y_i}{\Sigma x_i} - \frac{n \Sigma x_i y_i}{\Sigma x_i} + \frac{mn \Sigma x_i^2}{\Sigma x_i}$$

$$m\left(\frac{\Sigma x_i \Sigma x_i}{\Sigma x_i} - \frac{n \Sigma x_i^2}{\Sigma x_i}\right) = \frac{\Sigma x_i \Sigma y_i}{\Sigma x_i} - \frac{n \Sigma x_i y_i}{\Sigma x_i}$$

$$\frac{m}{\Sigma x_i}\left(\Sigma x_i \Sigma x_i - n \Sigma x_i^2\right) = \frac{1}{\Sigma x_i}\left(\Sigma x_i \Sigma y_i - n \Sigma x_i y_i\right)$$

$$m = \frac{\sum x_i \sum y_i - n \sum x_i y_i}{\sum x_i \sum x_i - n \sum x_i^2}$$

$$m = \frac{\sum x_i \sum y_i - n \sum x_i y_i}{(\sum x_i)^2 - n \sum x_i^2}$$

multiply num & denom by $-1$ each

$$m = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \quad \checkmark$$

---

Now solve for the y-intercept, $b$, using the equation $\frac{\partial Q}{\partial b} = 0$ ...

$$b = \frac{\Sigma y_i - m \Sigma x_i}{n}$$

$$b = \frac{\Sigma y_i}{n} - m \frac{\Sigma x_i}{n} \quad \checkmark$$

$$\left( b = \bar{y} - m \bar{x} \right)$$