

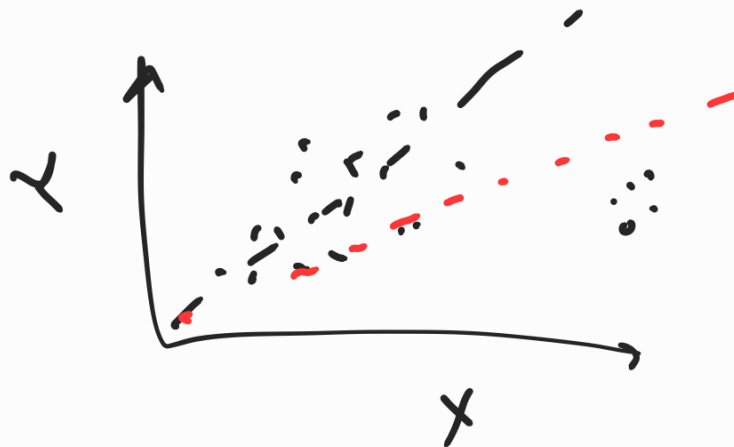
# CORRELATION

$$\text{corr}(X \sim Y) = R$$

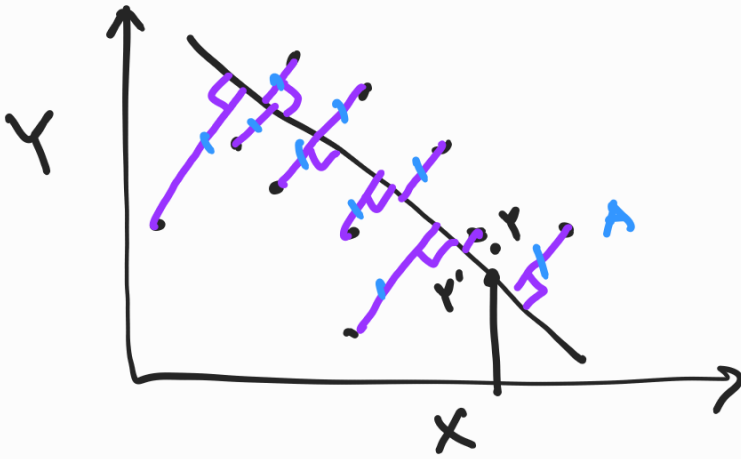
$+$ ,  $-$

$$-1 \leq R \leq 1$$

$R = 0$ , there is no relationship



# RESIDUALS



→ leftover distance.

→ data = model line  
+ residuals

→ residual =  $y - \hat{y}$   
residual error

# LEAST SQUARES LINE

$$f: x_i \rightarrow \hat{y}_i, y_i$$

$$\text{residual error } e_i = y_i - \hat{y}_i$$

Absolute error  $|e_1| + |e_2| + |e_3| + \dots + |e_n| \quad \left( \frac{L_1}{\text{norm}} \right)$

Squared error  $e_1^2 + e_2^2 + e_3^2 + \dots + e_n^2 \quad \left( \frac{L_2}{\text{norm}} \right)$

# GENERAL FORM OF LEAST SQUARES LINE

$$f: x_i \rightarrow \hat{y}_i$$

$$\hat{y}_i = \beta_0 + \beta_1 x_i$$
$$y = wx + c$$

optimize  
residual error

$$\sum (\hat{y}_i - y_i)$$

$$\text{for } \hat{y}_i = \beta_0 + \beta_1 x_i$$