Linear Regression

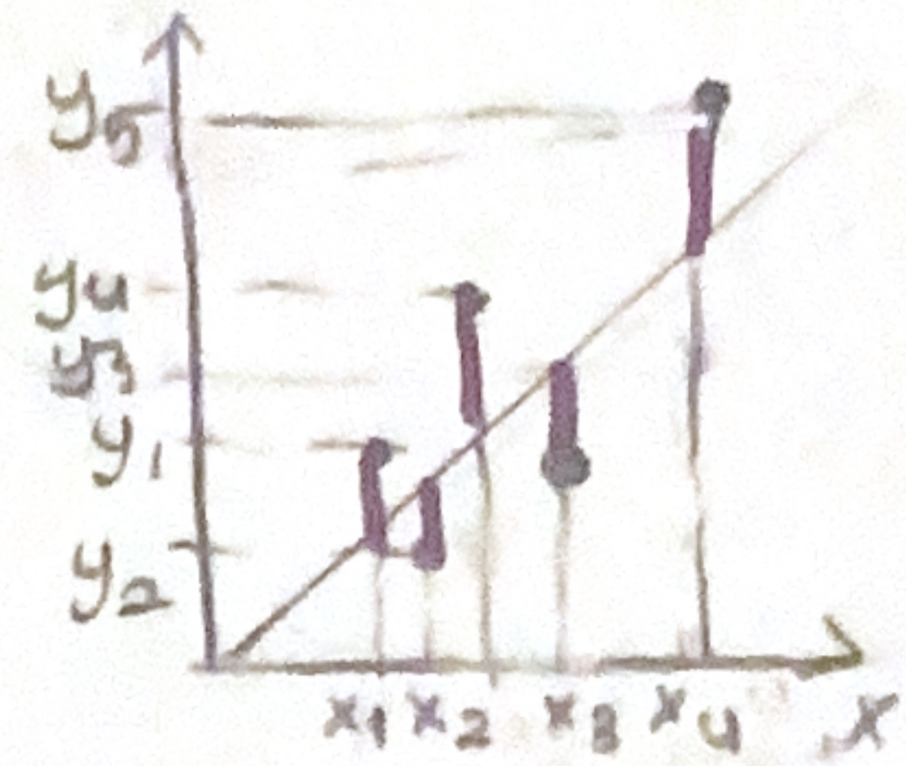
N data points

(x_1, y_1)
(x_2, y_2)
$\vdots$
(x_N, y_N)

$$y = \beta_0 + \beta_1 x_i + \underbrace{\varepsilon_i}_{\text{error term}}$$

$E[\varepsilon_i] = 0$
$V[\varepsilon_i] = \sigma^2$

Assume it's a RV, distributed normally.



↳ least squares

min $\sum \varepsilon_i^2$ by changing β_0 & β_1

* $$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$$

$$\boxed{\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}}$$

* $$S_{xy} = \sum x_i y_i - \frac{\sum x_i * \sum y_i}{N}$$

* $$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{N}$$

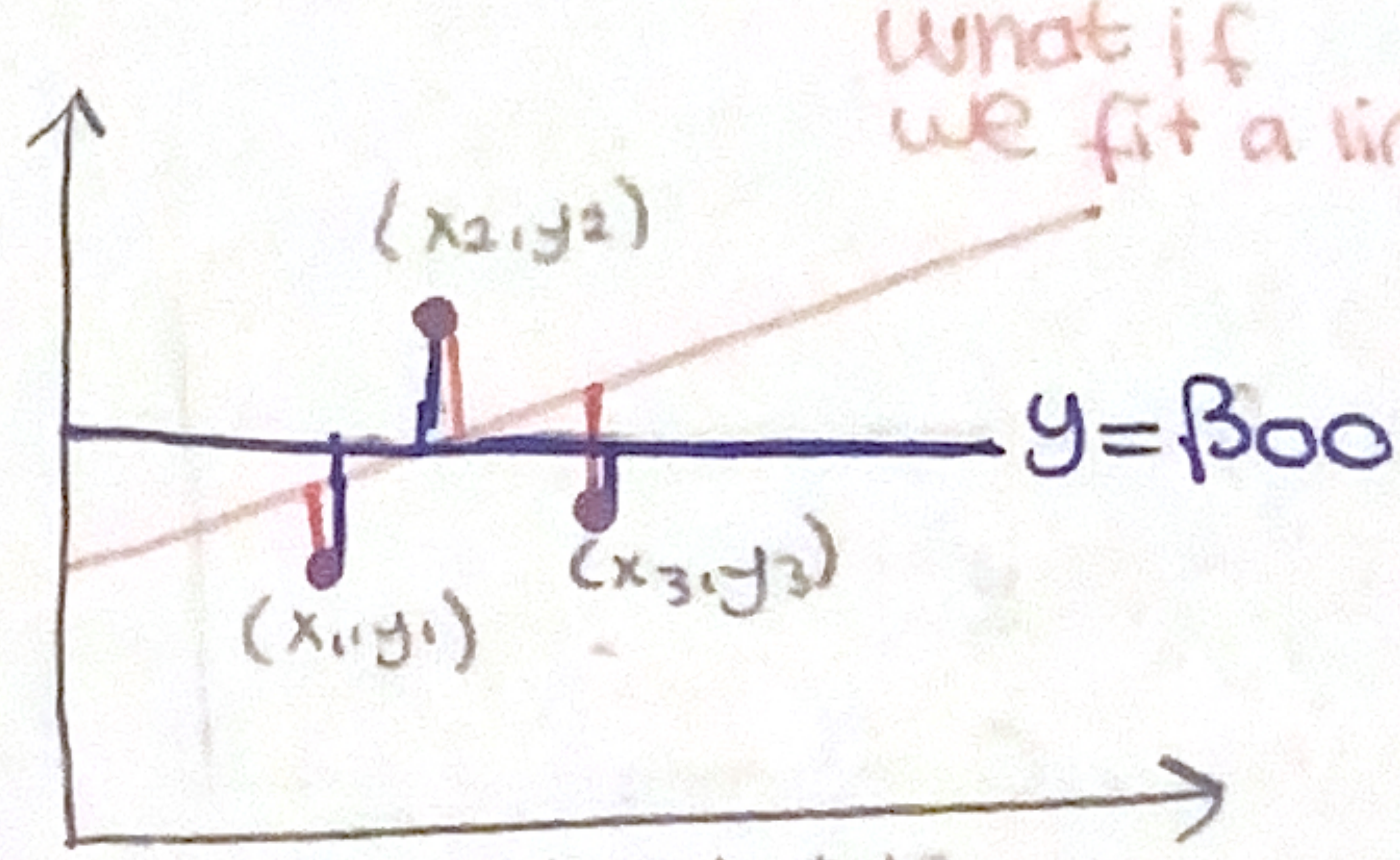
→ **Do the fit!**

Goodness of Fit

$y = \beta_0 + \beta_1 x$ → Our fit

$y = \beta_{00}$ → what else could be done?

what if we fit a line?



$$y = \bar{y} = \frac{y_1 + y_2 + y_3}{3}$$

* $SST = \sum (y_i - \bar{y})^2$
↓
sum of squares of total

* $SSE = \sum (y_i - \beta_0 - \beta_1 x_i)^2$ → $0 \leq \frac{SSE}{SST} \leq 1$

$\frac{SSE}{SST} = 1$ ↳ Bad fit

↳ Bunun noktası daha açık olmalı

$$\boxed{R^2 = 1 - \frac{SSE}{SST}}$$