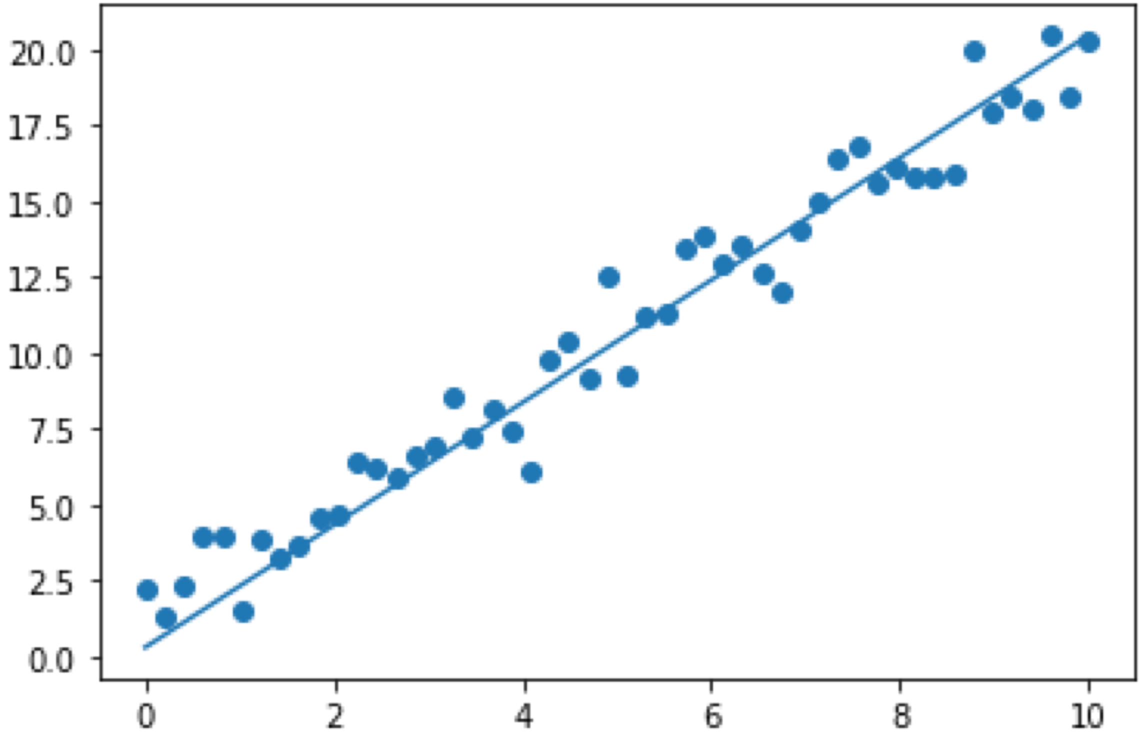
Linear Regression In Depth



Simple Linear Regression (1-D, vector form)

- You have a dataset: $\{(x_i, y_i)\}_{i \in [1,2,\ldots n]}$
- You wish to find the "best" curve relating your target, y, to your features, x.

- Assume: $y_i = E[y_i] + \varepsilon_i,\ \text{(Mean trend + Noise)}$
- Where, $E[y_i] = \beta_0 + \beta_1 x_i,$
  $\varepsilon \sim N(0, \sigma^2)$

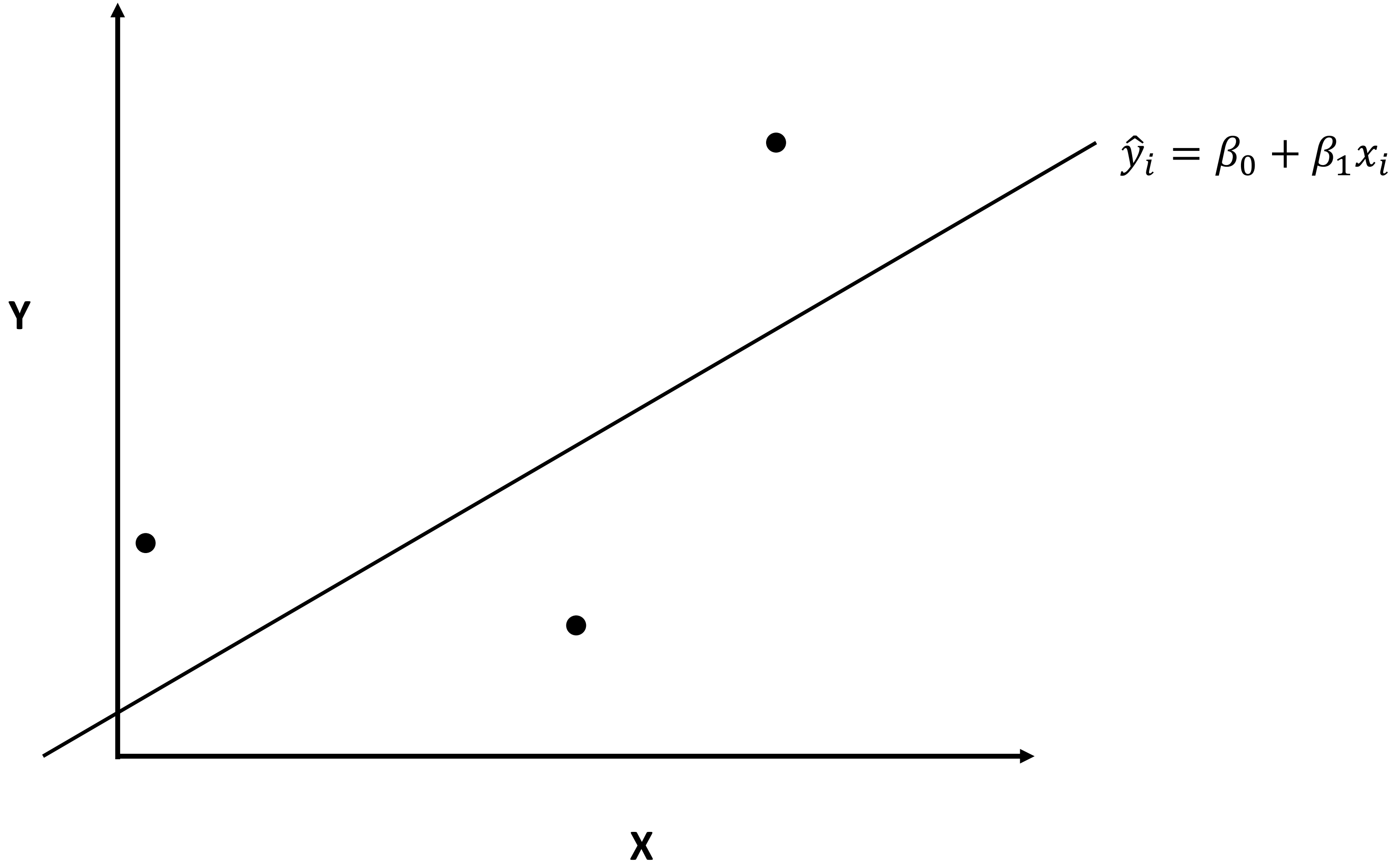
- Equivalently: $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$

Simple Linear Regression (1-D)

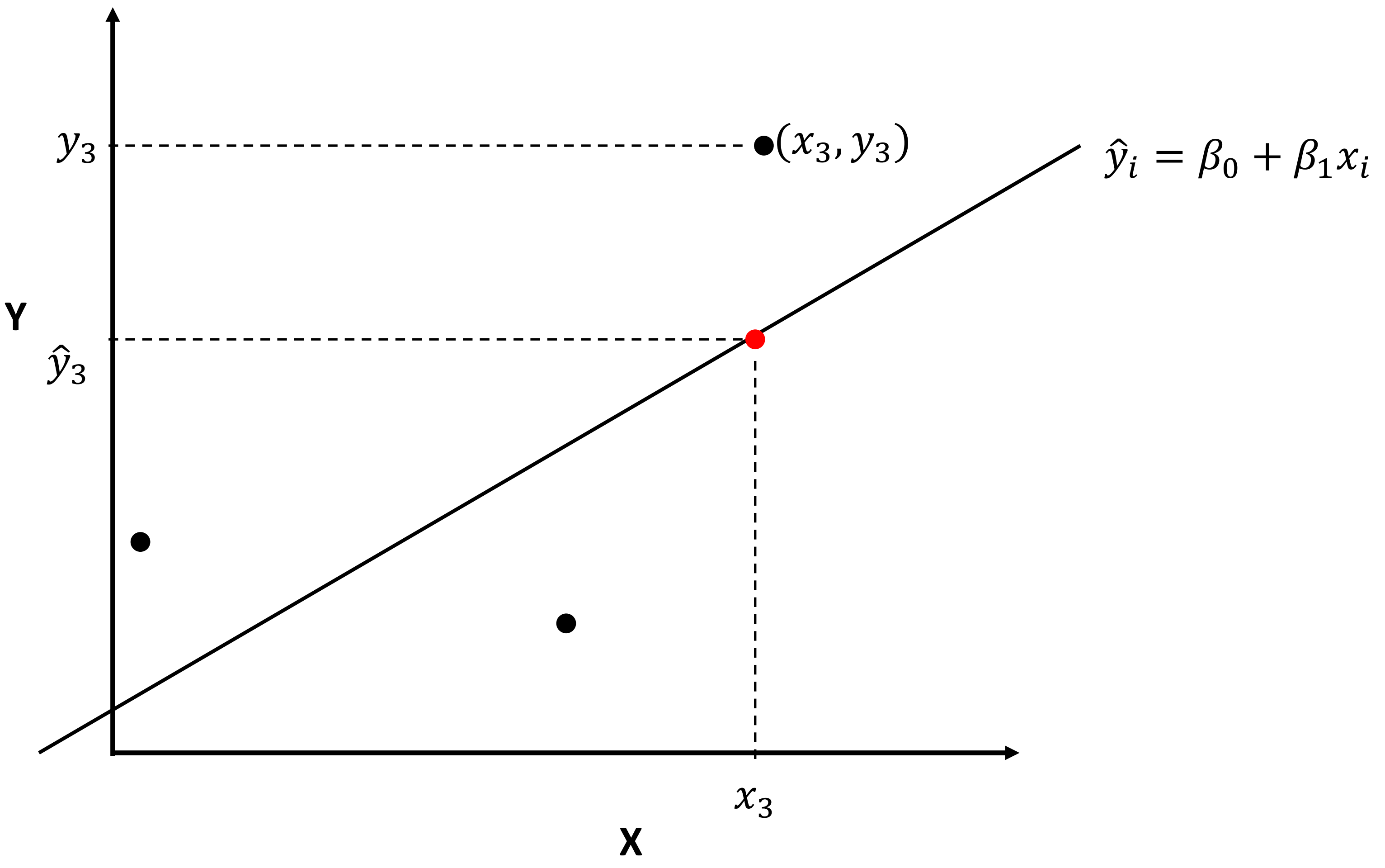
- Underlying Assumptions: **L**-**I**-**N**-**E**,

- L: Regression is Linear,
- I: Errors are Independent,
- N: Errors are Normally distributed,
- E: Errors have constant variance.

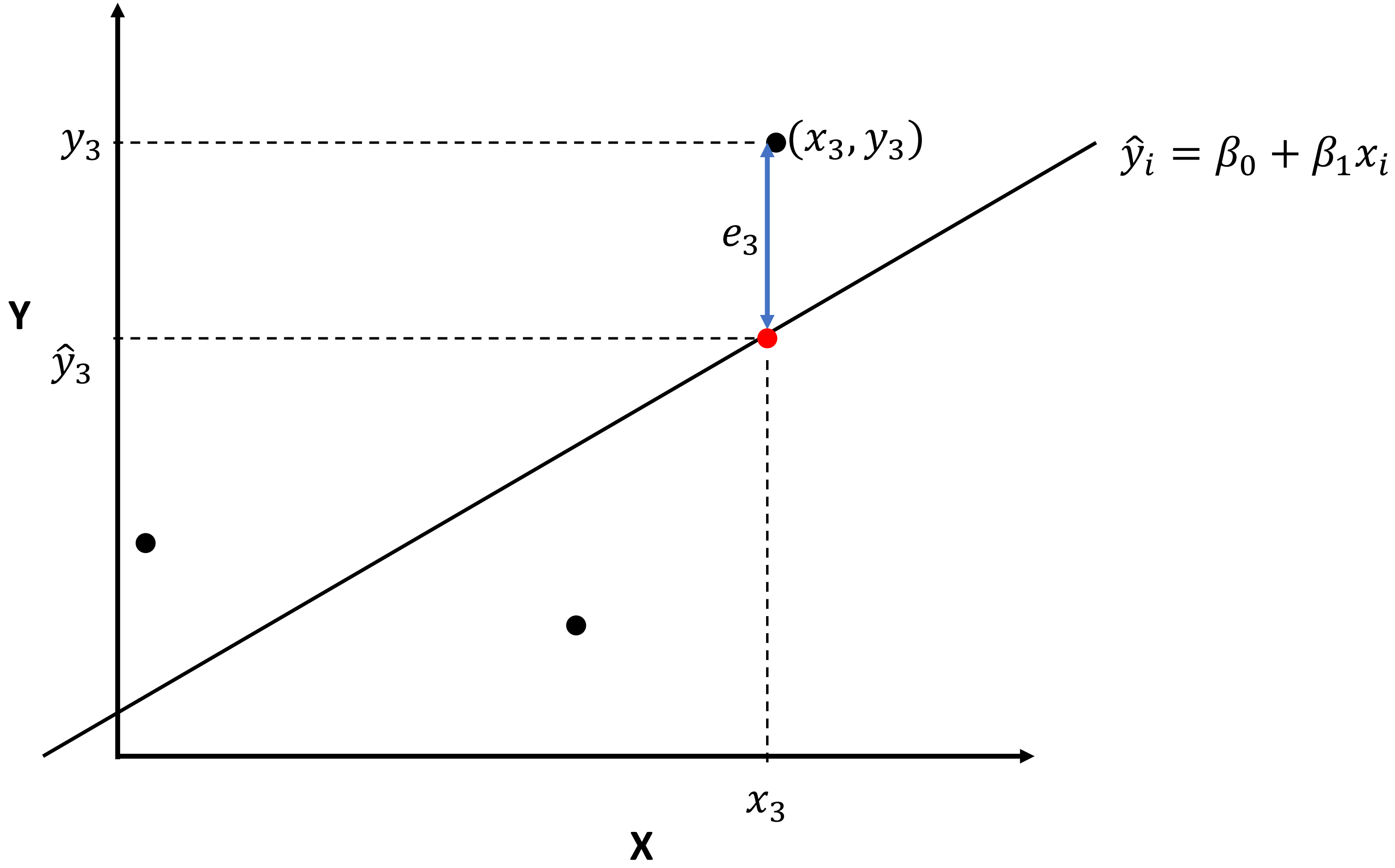
Simple Linear Regression (1-D)



Simple Linear Regression (1-D)

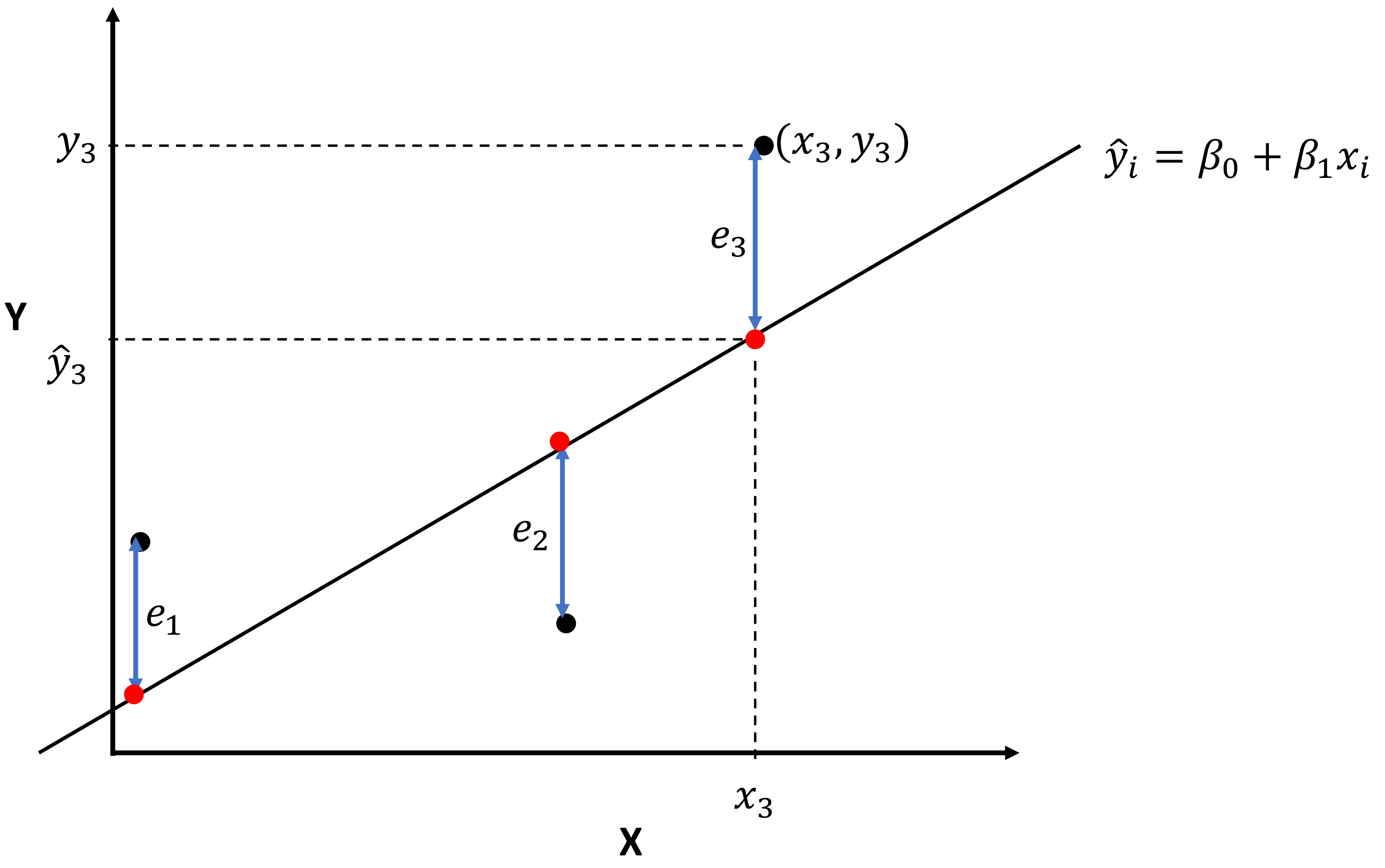


Simple Linear Regression (1-D)



$e_3 = |y_3 - \hat{y}_3|$
$e_3 = |y_3 - (\beta_0 + \beta_1 x_3)|$

Simple Linear Regression (1-D)



Simple Linear Regression (1-D)

- Mean Squared Error: $E = \frac{1}{n}\sum_{i=1}^{n} e_i^2$
- But $e_i = [y_i - (\beta_0 + \beta_1 x_i)]$
- Hence, $E(\beta_0, \beta_1) = \frac{1}{n}\sum_{i=1}^{n}[y_i - (\beta_0 + \beta_1 x_i)]^2$

Simple Linear Regression (1-D)

- To perform Gradient Descent, we need the gradient:

$$\begin{bmatrix} \frac{\partial E}{\partial \beta_0} \\ \frac{\partial E}{\partial \beta_1} \end{bmatrix} = \frac{1}{n} \sum_{i=1}^{n} (-2) \begin{bmatrix} [y_i - (\beta_0 + \beta_1 x_i)] \\ x_i [y_i - (\beta_0 + \beta_1 x_i)] \end{bmatrix}$$

- And for every iteration of gradient descent:

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} - \alpha \begin{bmatrix} \frac{\partial E}{\partial \beta_0} \\ \frac{\partial E}{\partial \beta_1} \end{bmatrix}$$

Simple Linear Regression (Matrix Form)

- Design Matrix: $X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ .. & .. \\ 1 & x_n \end{bmatrix}$
- Weights Vector: $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$
- Model Prediction: $\hat{Y} = X\beta$

Simple Linear Regression (Matrix Form)

- Mean Squared Error: $E = \left(Y - \hat{Y}\right)^T (Y - \hat{Y})/n$

- Gradient: $\frac{\partial E}{\partial \beta} = (\beta^T X^T X \ - \ Y^T Y)^T/n$