CIS 678 Machine Learning

ML Introduction: Linear Regression (part 2)

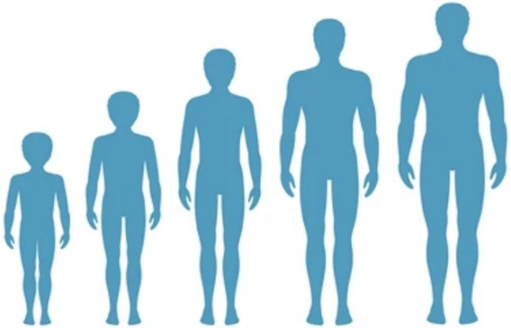
What we'd like to accomplish today

- ~~General Concepts: ***Straight Line*** *to* ***Linear Regression***~~
- Gradient Descent Algorithm
    - ➜ A simple two parameter **Linear Regression** model
    - ➜ Hands on **Notebook implementation**
- QA

We will be learning today about Regression

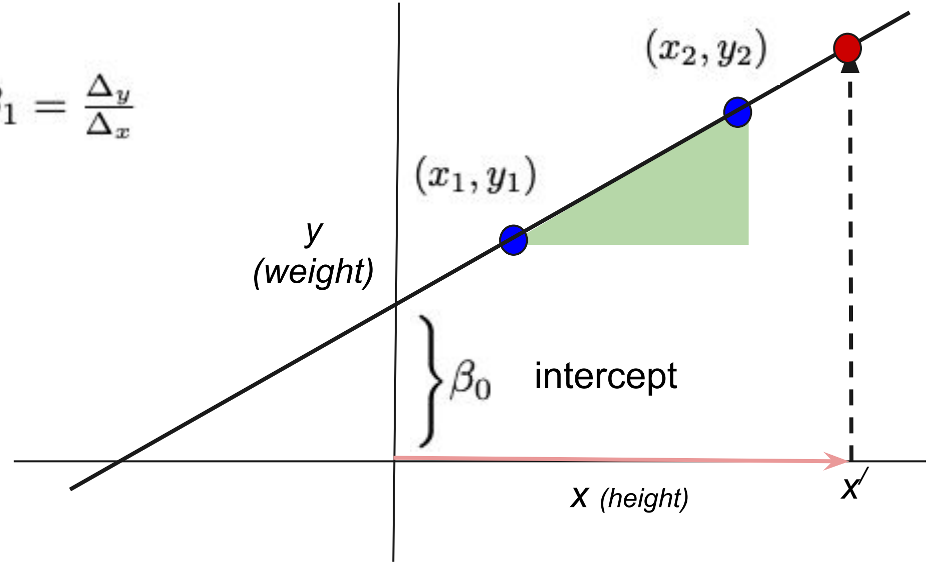
- *Person's weight :* $y \in R$

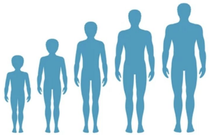
$$f\,(y|x=height)$$



Linear equation to a linear function, a quick review

slope $\beta_1 = \frac{\Delta_y}{\Delta_x}$





Was able to capture the LR relationship

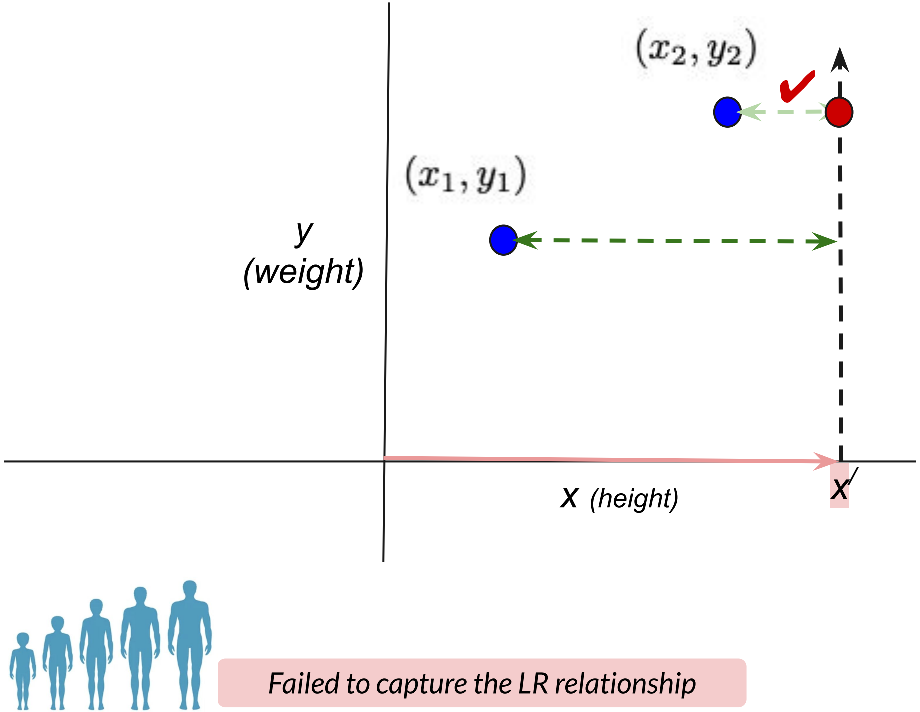
- This linear equation can be used to explain the relationship between the two axes ***x*** (independent variable) vs ***y*** (dependent variable) - as

$$y = \beta_0 + \beta_1 x$$

- A simple model with parameters: **slope**, and **intercept**

- For any given x', this model can predict $y(x')$ using the above equation.

k-NN Regression



Failed to capture the LR relationship

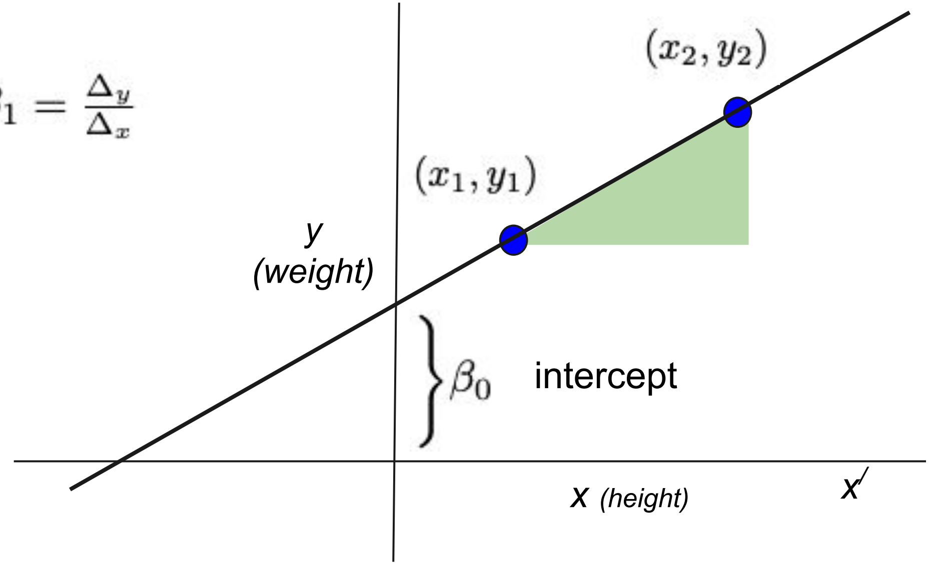
Given two known data points (x_1,y_1), and (x_2,y_2), and

- for test input x', you have to predict $y(x')$.
- I.e. you have to plot (x', ?)
- To estimate the distances let's draw the vertical line
- Horizontal dotted lines show the point distances (L1)
- We find the lighter green on is the closest one [k(1)-NN]
- We propagate the associated label(s), i.e.

  $y(x') = y_2$
- If we have more data points we may go for a higher k, and take the average

Linear equation to a linear function, a quick review

slope $\beta_1 = \frac{\Delta_y}{\Delta_x}$



Linear line model, θ: {β_0, β_1}

- This linear equation can be used to explain the relationship between the two axes ***x*** (independent variable) vs ***y*** (dependent variable) - as

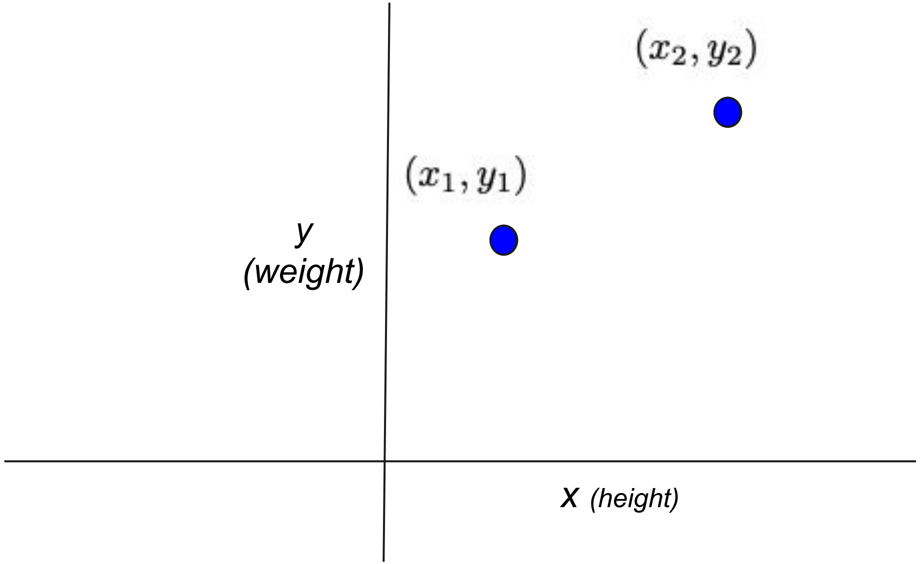
$$y = \beta_0 + \beta_1 x$$

- A simple model with parameters: **slope**, and **intercept**

- For any given x', this model can predict $y(x')$ using the above equation.

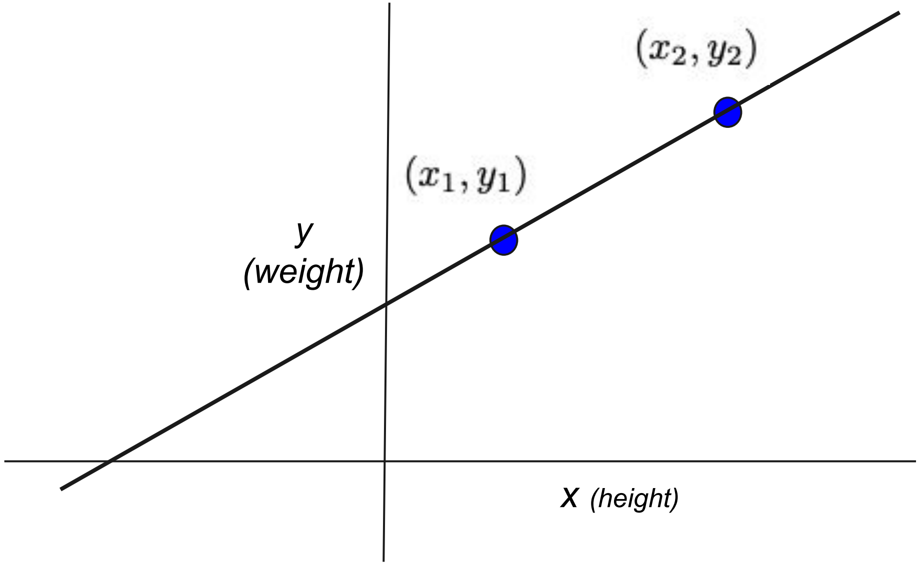
From Linear Equation to Linear Regression

Linear equation, a quick review



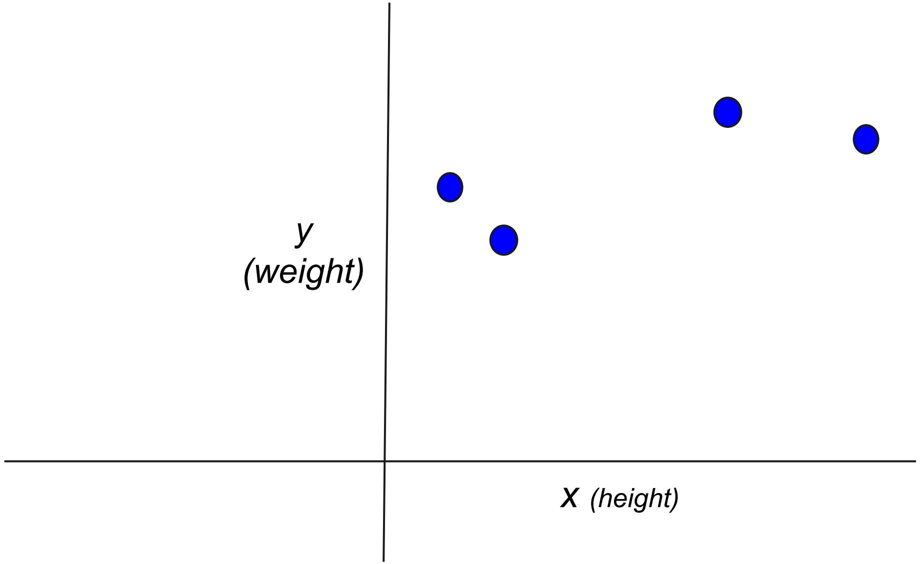
- *Given two (2) data points,*
- *We can fit a linear line.*
- *What if we have > 2 data points?*

Linear equation, a quick review



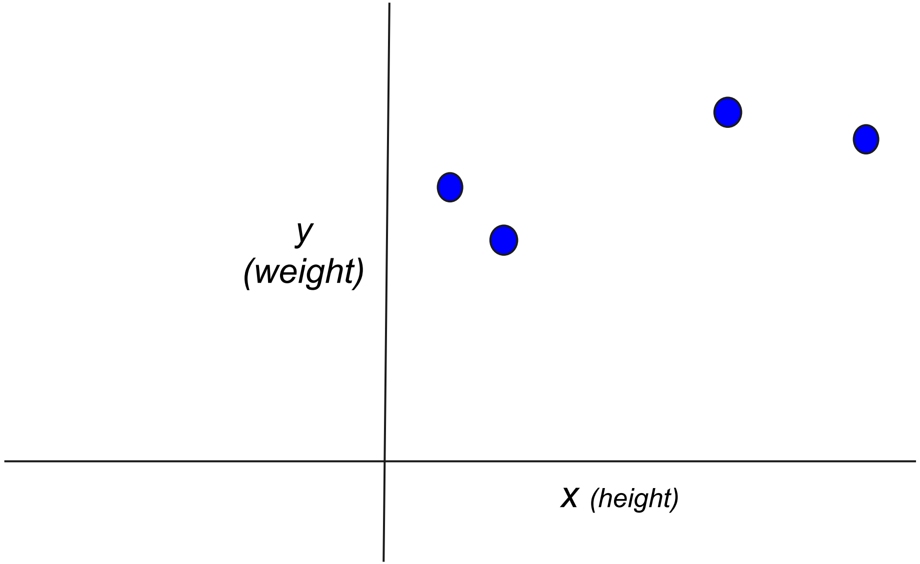
- Given two (2) data points,
- We can fit a linear line.

Straight Line to Linear Regression



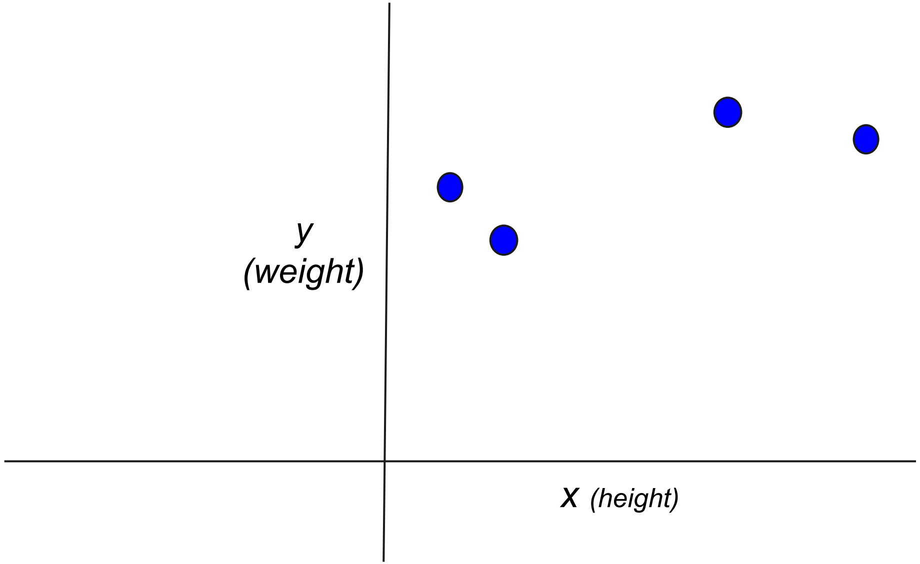
- What if we have > 2 data points?
- It is very unlikely that new points will fit on the same straight line.
- We cannot fit them through a linear line.
- We can try many (in fact infinite) ways.

Straight Line to Linear Regression



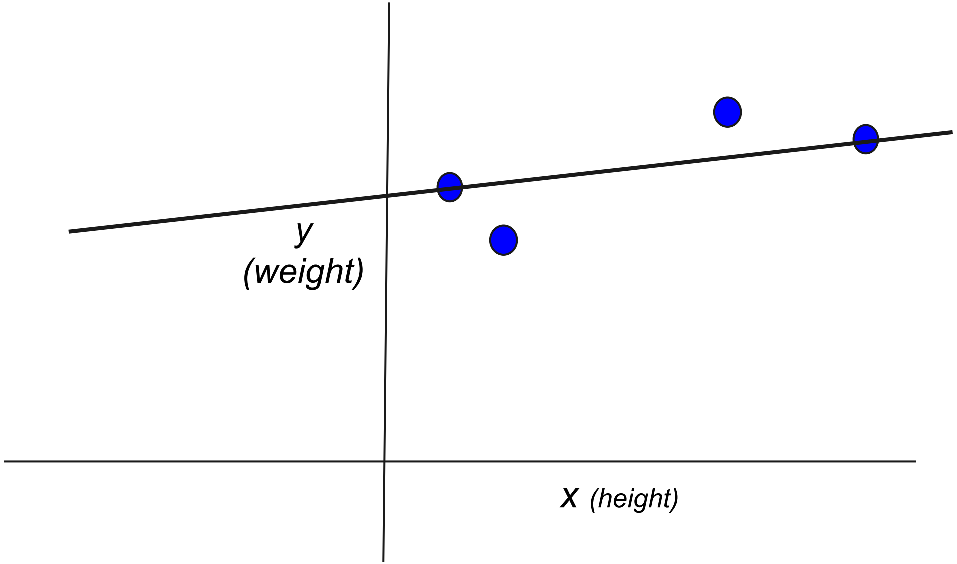
- What if we have > 2 data points?
- It is very unlikely that new points will fit on the same straight line.
- We cannot fit them through a linear line.
- We can try many (in fact infinite) ways.

Straight Line to Linear Regression



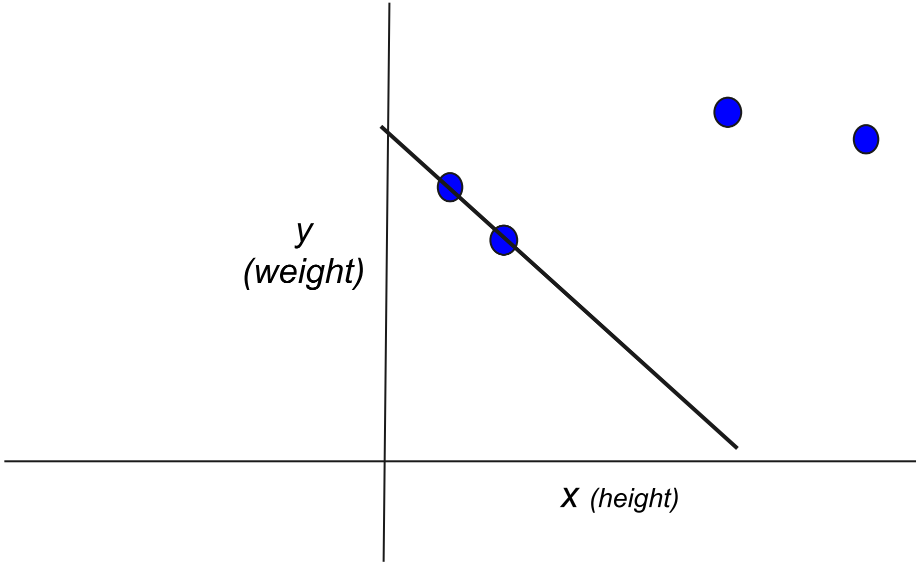
- What if we have > 2 data points?
- It is very unlikely that new points will fit on the same straight line.
- We cannot fit them through a linear line.
- We can try many (in fact infinite) ways.

Straight Line to Linear Regression



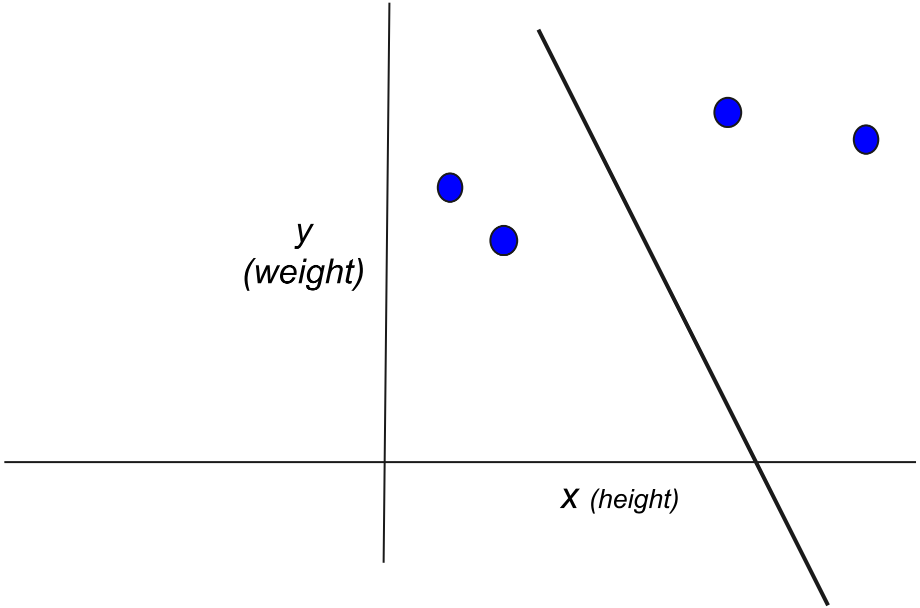
- What if we have > 2 data points?
- It is very unlikely that new points will fit on the same straight line.
- We cannot fit them through a linear line.
- We can try many (in fact infinite) ways.

Straight Line to Linear Regression



- What if we have > 2 data points?
- It is very unlikely that new points will fit on the same straight line.
- We cannot fit them through a linear line.
- We can try many (in fact infinite) ways.

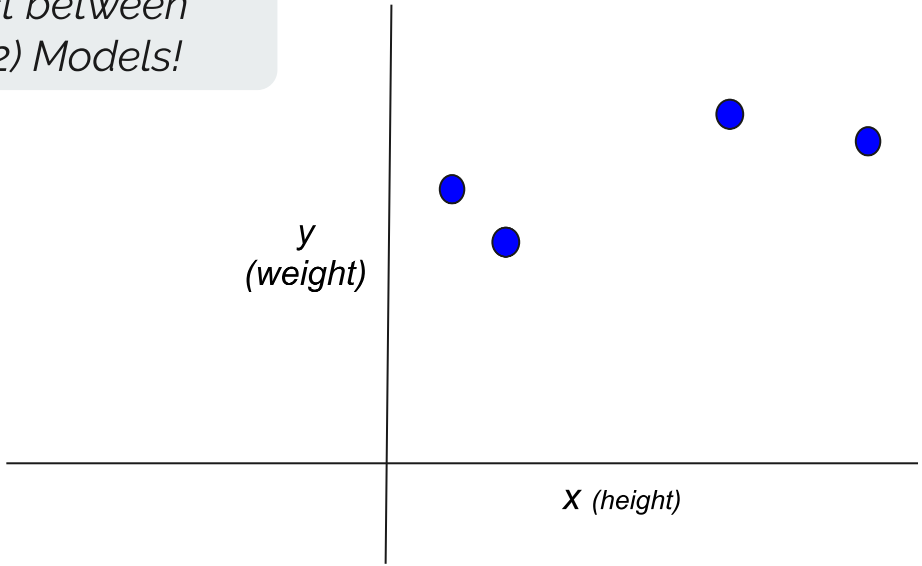
Straight Line to Linear Regression



- *What if we have > 2 data points?*
- *It is very unlikely that new points will fit on the same straight line.*
- *We cannot fit them through a linear line.*
- *We can try many (in fact infinite) ways.*

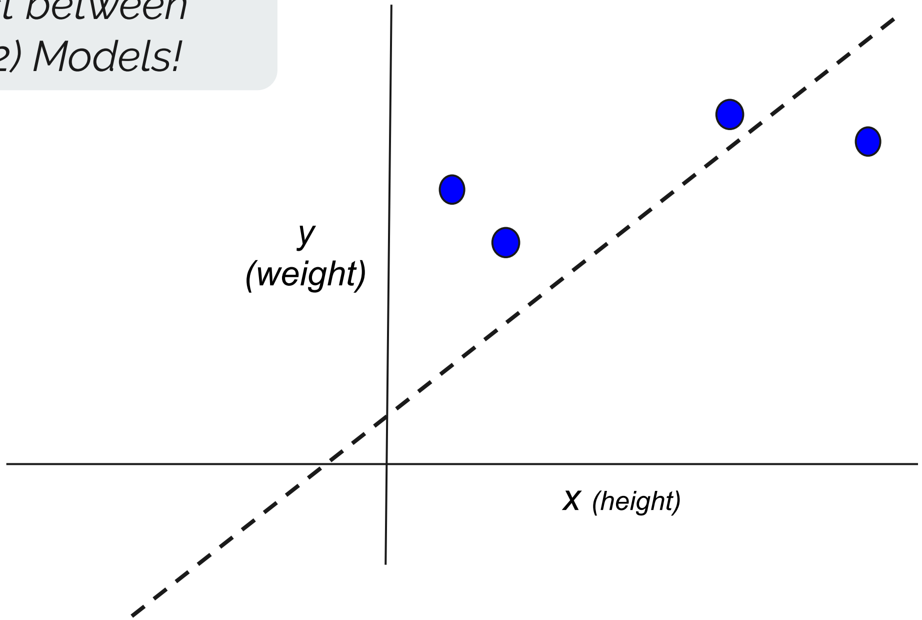
Straight Line to Linear Regression

Select between two (2) Models!



Straight Line to Linear Regression

Select between two (2) Models!



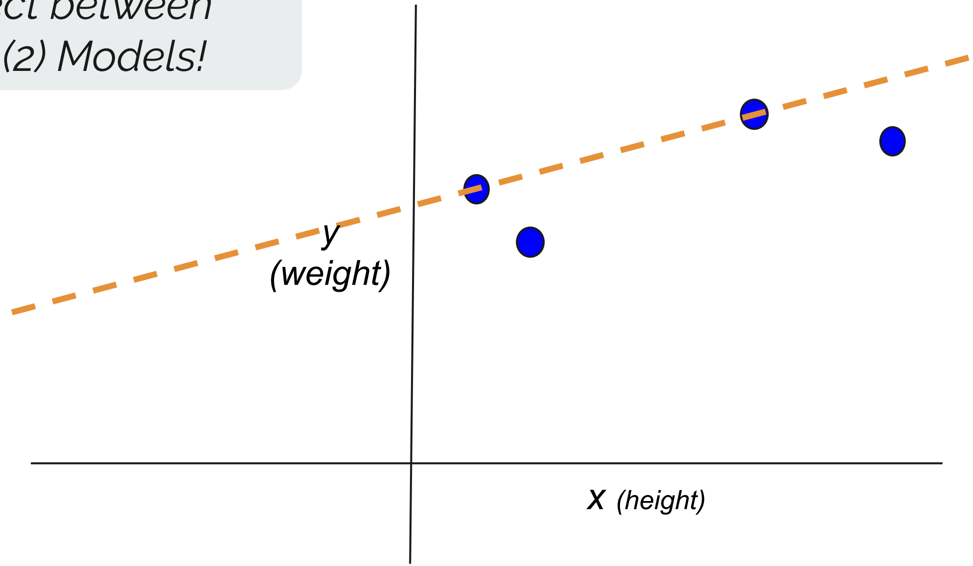
- *Given the following 2 models*

Model, θ: {β_0, β_1}

- *Which one will you suggest?*

Straight Line to Linear Regression

Select between two (2) Models!



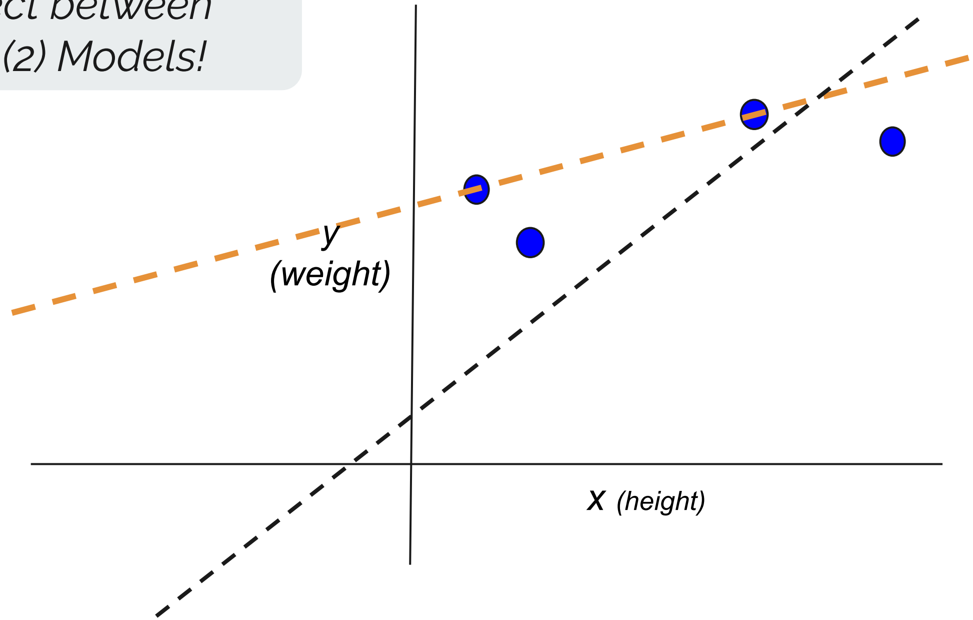
- *Given the following 2 models*

Model, θ: {β_0, β_1}

- *Which one will you suggest?*

Straight Line to Linear Regression

Select between two (2) Models!



- *Given the following 2 models*

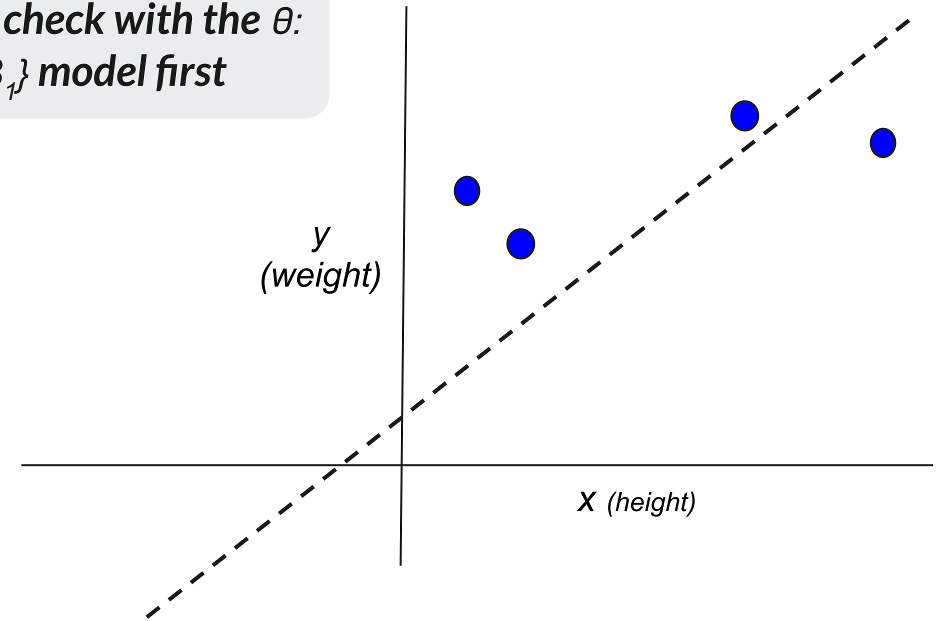
Model, θ: {β_0, β_1}

Model, θ: {β_0, β_1}

- *Which one will you suggest?*

Straight Line to Linear Regression

Let's check with the θ: $\{\beta_0, \beta_1\}$ model first



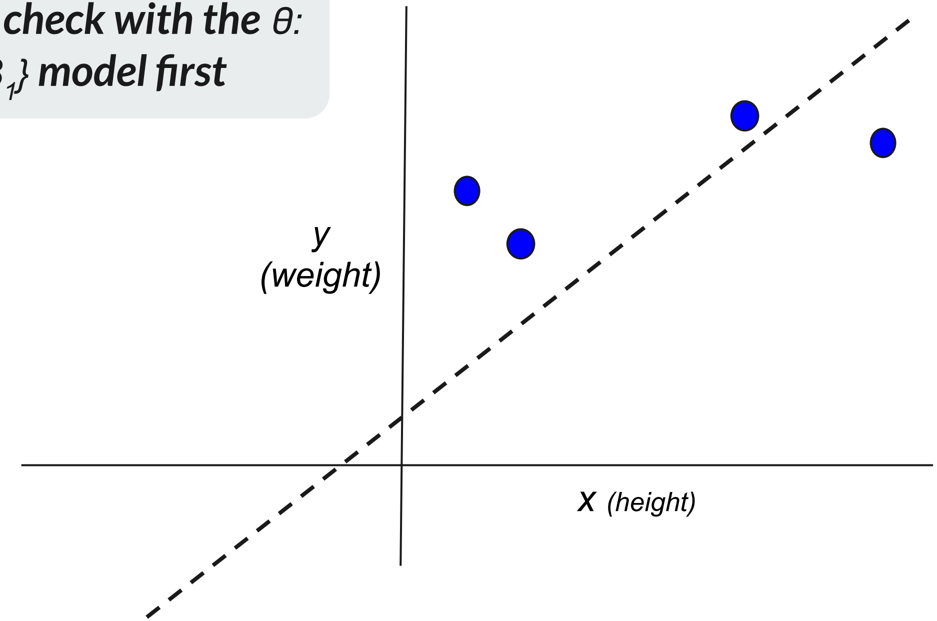
- *Given the following 2 models*

Model, θ: $\{\beta_0, \beta_1\}$

- *Which one will you suggest?*
- *We require some strategies, say apply some performance quantification method.*

Straight Line to Linear Regression

Let's check with the θ: $\{\beta_0, \beta_1\}$ model first



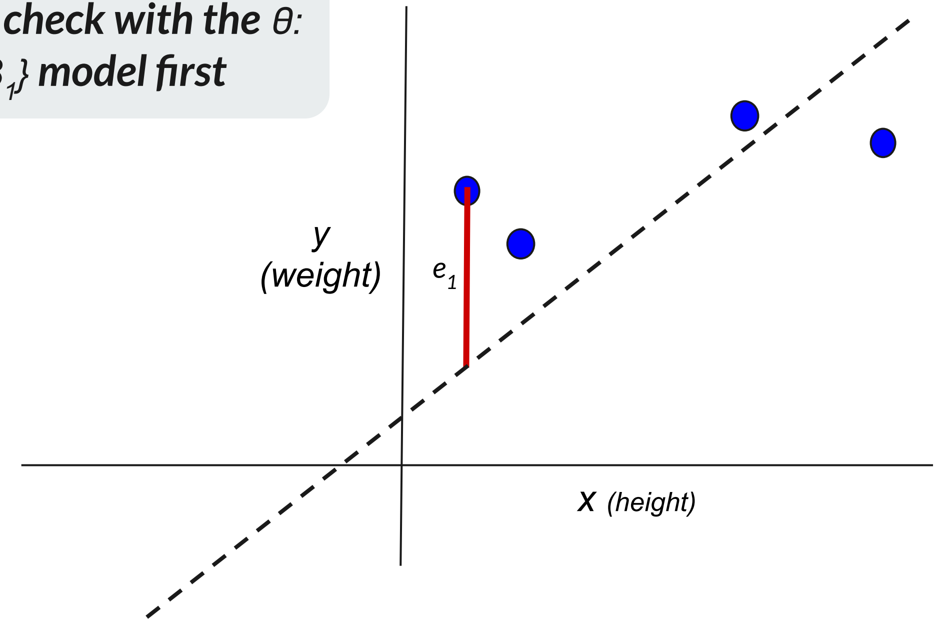
- *Given the following 2 models*

Model, θ: $\{\beta_0, \beta_1\}$

- *Which one will you suggest?*
- *We require some strategies, say apply some performance quantification method.*
- *We can use Absolute error as our model performance metric.*

Straight Line to Linear Regression

Let's check with the θ: $\{\beta_0, \beta_1\}$ model first



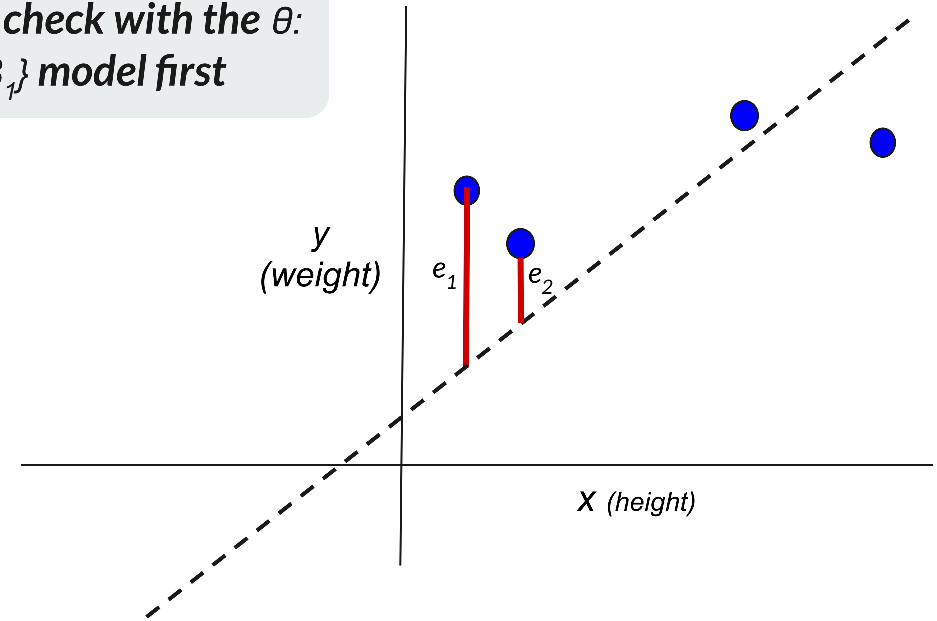
- *Given the following 2 models*

Model, θ: $\{\beta_0, \beta_1\}$

- *Which one will you suggest?*
- *We require some strategies, say apply some performance quantification method.*
- *We can use Absolute error as our model performance metric.*

Straight Line to Linear Regression

Let's check with the θ: $\{\beta_0, \beta_1\}$ model first



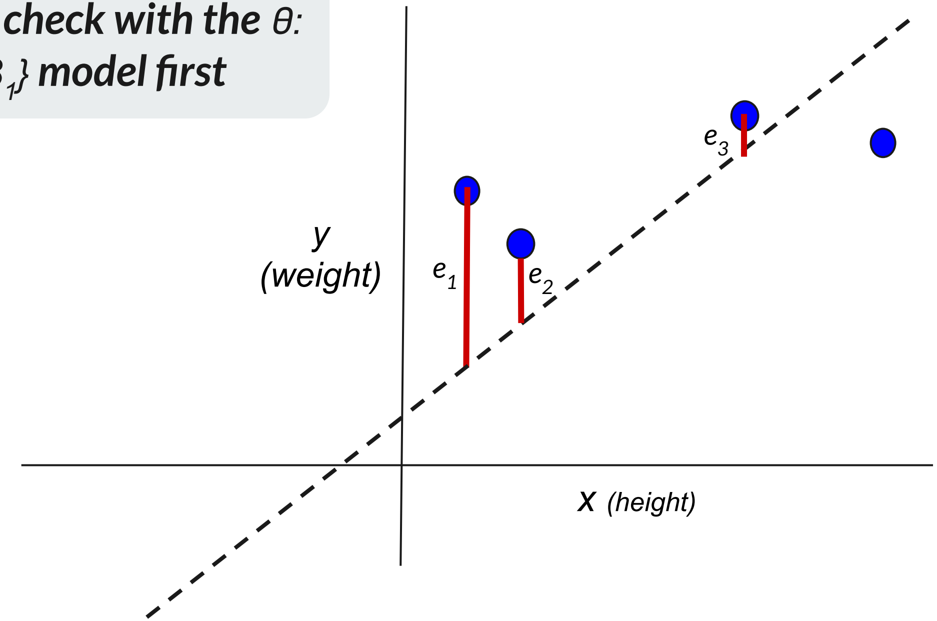
- *Given the following 2 models*

Model, θ: $\{\beta_0, \beta_1\}$

- *Which one will you suggest?*
- *We require some strategies, say apply some performance quantification method.*
- *We can use Absolute error as our model performance metric.*

Straight Line to Linear Regression

Let's check with the θ: $\{\beta_0, \beta_1\}$ model first



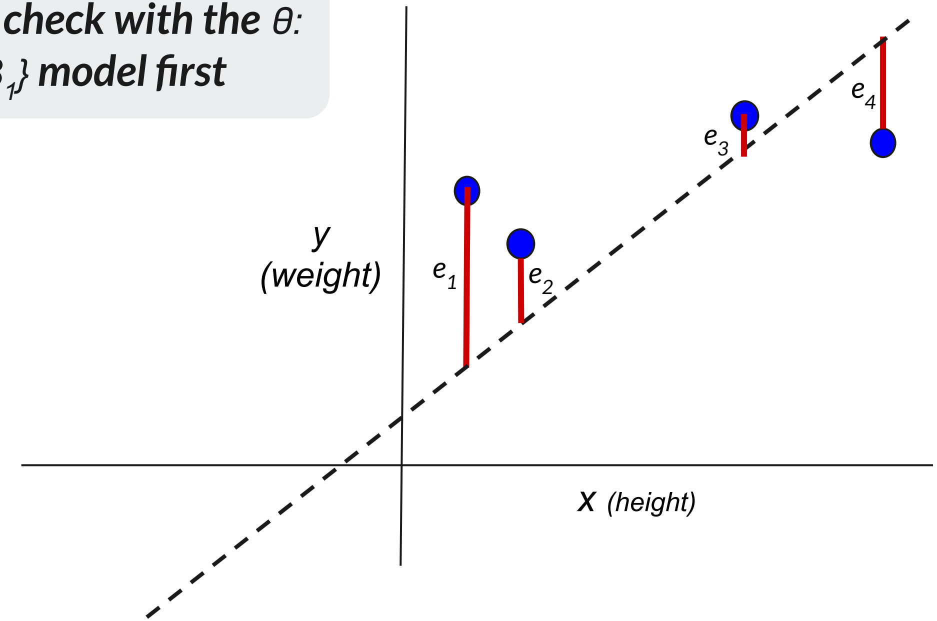
- *Given the following 2 models*

Model, θ: $\{\beta_0, \beta_1\}$

- *Which one will you suggest?*
- *We require some strategies, say apply some performance quantification method.*
- *We can use Absolute error as our model performance metric.*

Straight Line to Linear Regression

Let's check with the θ: {β_0, β_1} model first



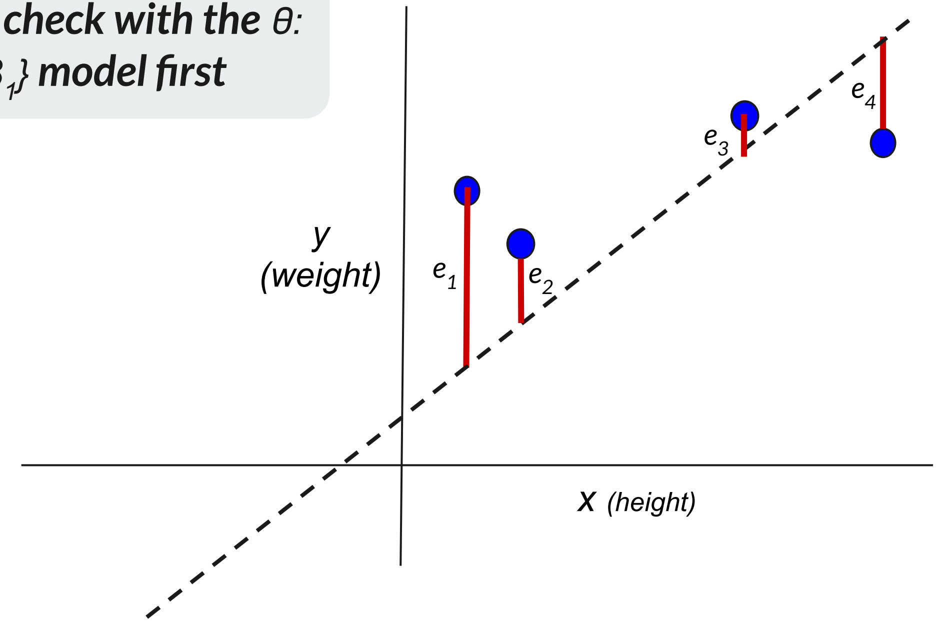
- *Given the following 2 models*

Model, θ: {β_0, β_1}

- *Which one will you suggest?*
- *We require some strategies, say apply some performance quantification method.*
- *We can use Absolute error as our model performance metric.*

Straight Line to Linear Regression

Let's check with the θ: $\{\beta_0, \beta_1\}$ model first



Total error: $e_1 + e_2 + e_3 + e_4$

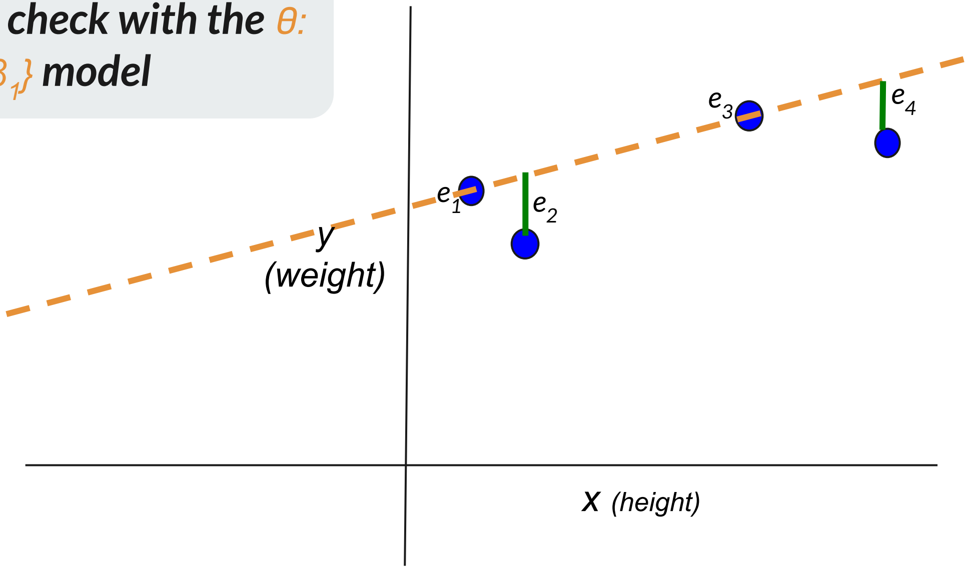
- *Given the following 2 models*

Model, θ: $\{\beta_0, \beta_1\}$

- *Which one will you suggest?*
- *We require some strategies, say apply some performance quantification method.*
- *We can use Absolute error as our model performance metric.*

Straight Line to Linear Regression

Let's check with the θ: $\{\beta_0, \beta_1\}$ ***model***



- *Given the following 2 models*

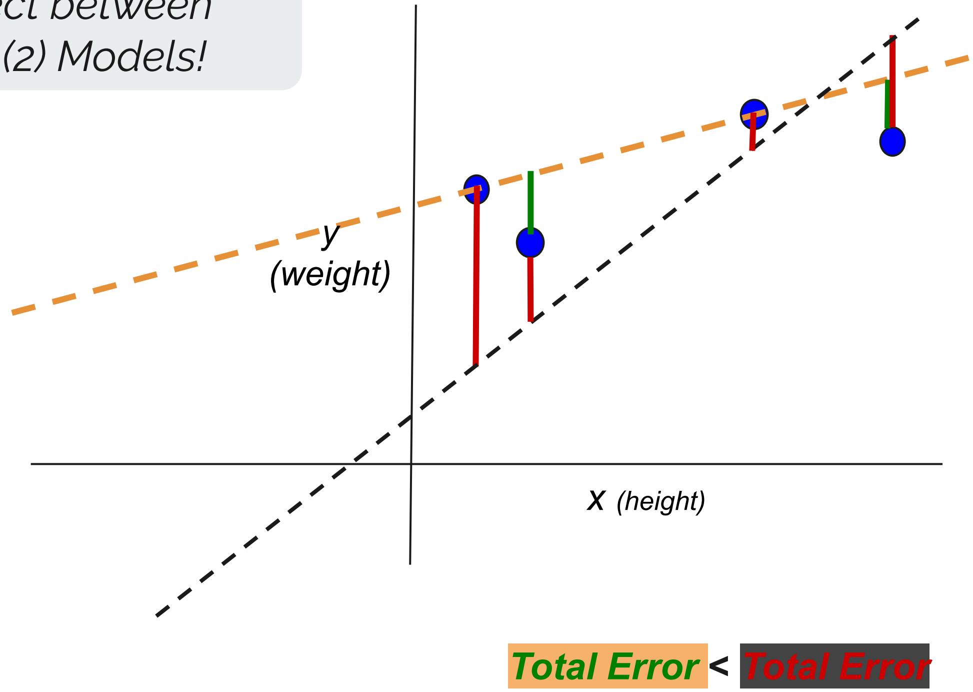
Model, θ: $\{\beta_0, \beta_1\}$

- *Which one will you suggest?*
- *We require some strategies, say apply some performance quantification method.*
- *We can use Absolute error as our model performance metric.*

Total error: $0 + e_2 + 0 + e_4$

Straight Line to Linear Regression

Select between two (2) Models!



- *Given the followi*

Model, θ:

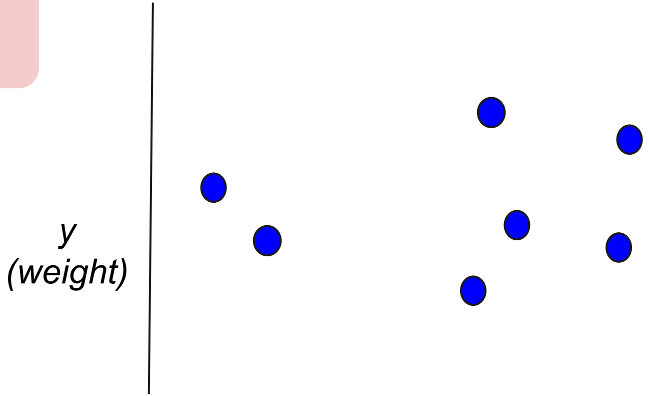
Model, θ: {β_0, β_1}

Our choice
(based on Abs error)

- *Which one will you suggest?*

Select one among a set of Models!

Any number of data points!



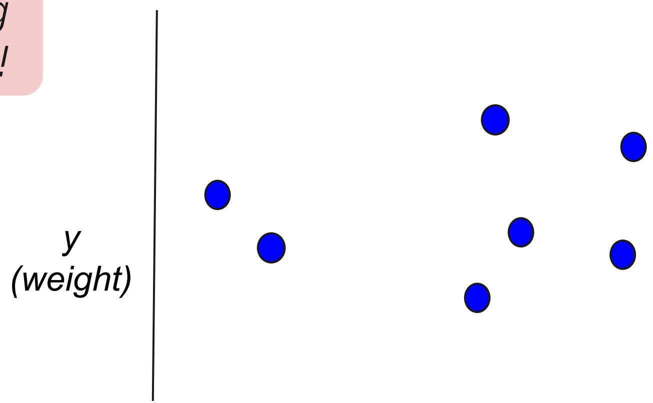
_________ ***How to Generalize the Idea for***

- any number of points, and/or

- any models (remember, we have infinite number of possible models)

Select one among a set of Models!

Select one among infinite (∞) Models!



Model, θ: $\{\beta_0, \beta_1\}$

Model, θ: $\{\beta_0, \beta_1\}$

Model, θ: $\{\beta_0, \beta_1\}$

Model, θ: $\{\beta_0, \beta_1\}$

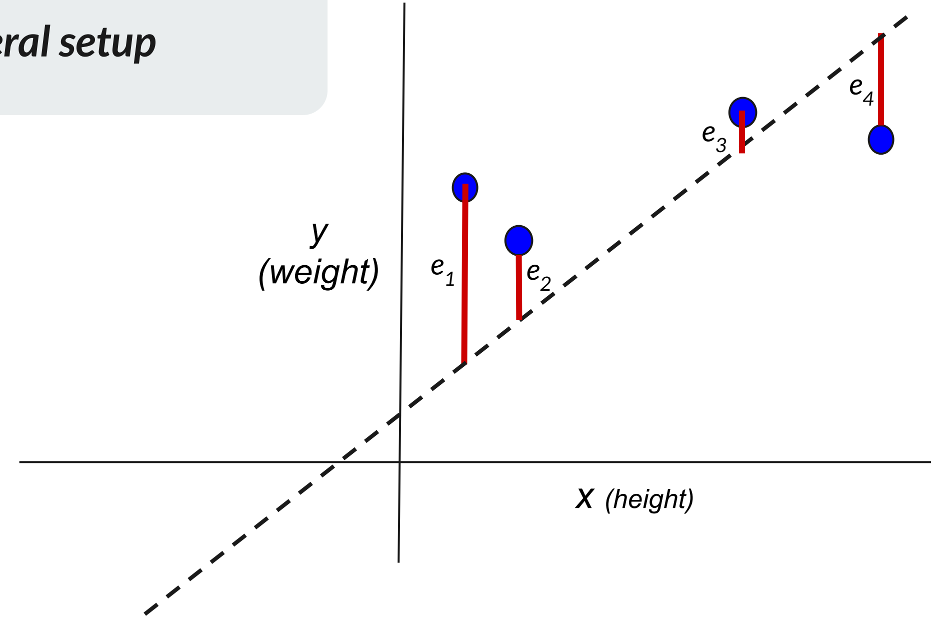
________ ***How to Generalize the Idea for***

- any number of points, and/or

- any models (remember, we have infinite number of possible models)

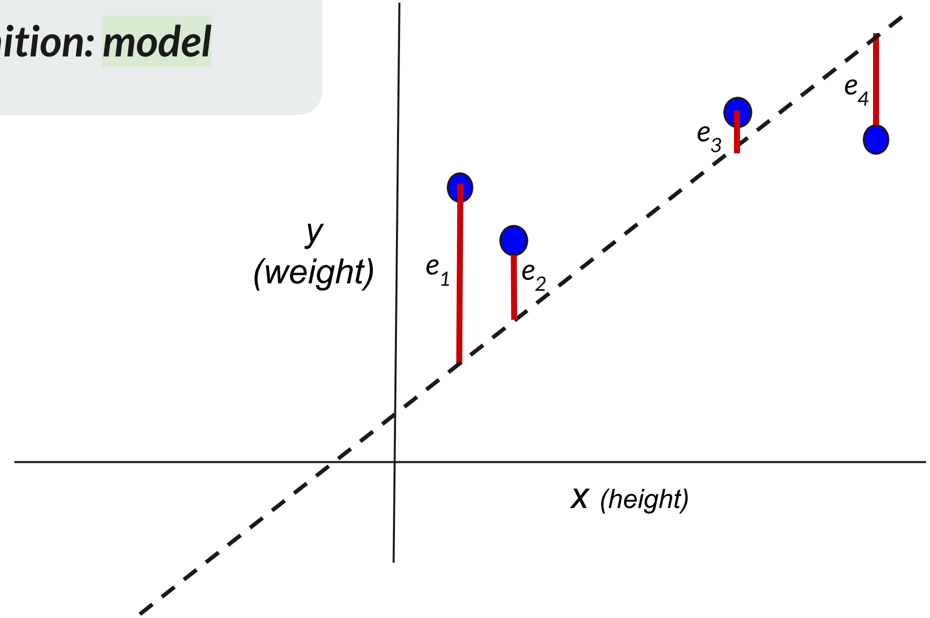
Linear Regression

General setup



Linear Regression

Definition: model



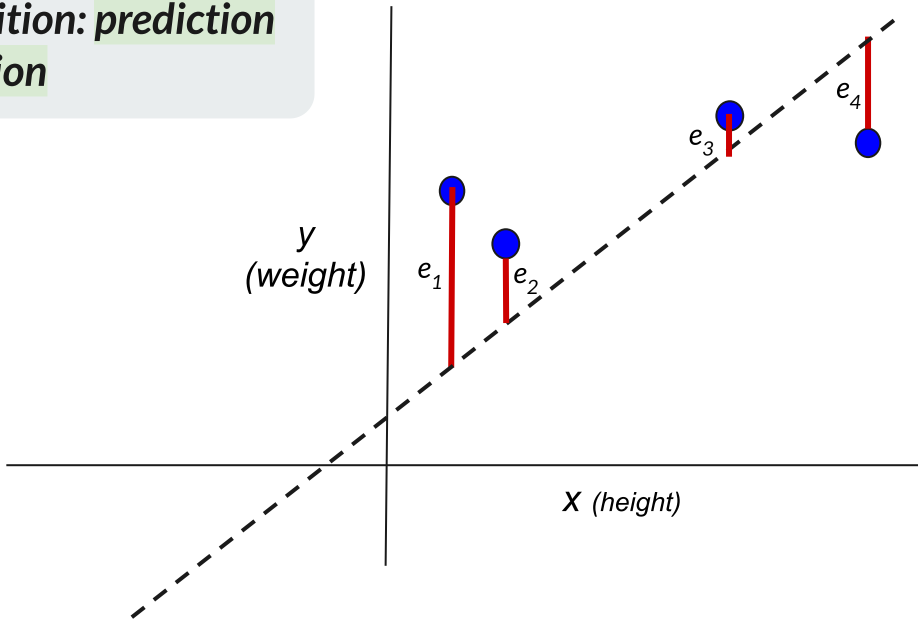
Prediction function & Model

$$\hat{y} = \beta_0 + \beta_1 x$$

$$\boxed{\Theta = \{\beta_0, \beta_1\}}$$

Linear Regression

Definition: prediction function



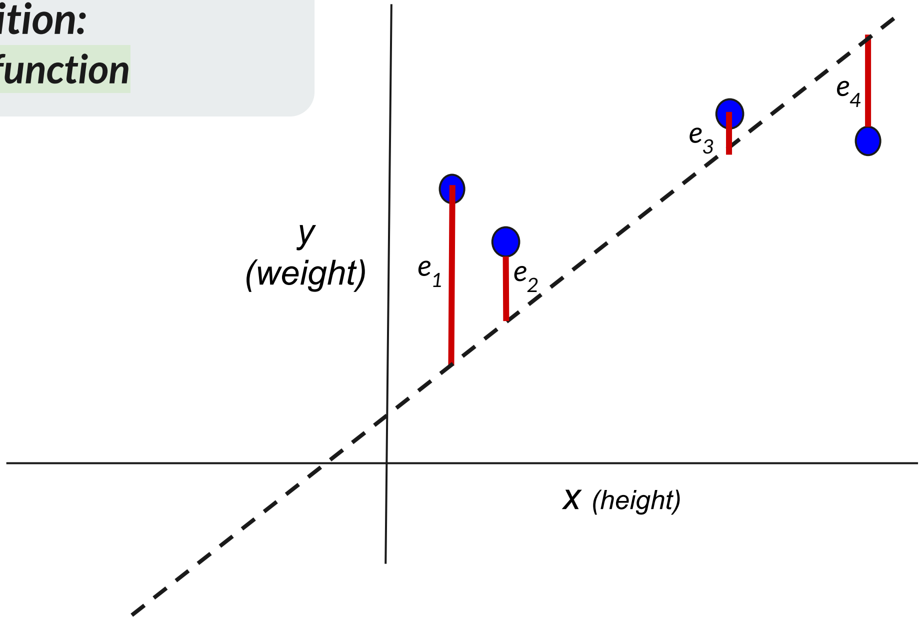
Prediction function & Model

$$\boxed{\hat{y} = \beta_0 + \beta_1 x}$$

$$\Theta = \{\beta_0, \beta_1\}$$

Linear Regression

Definition:
Error function



Prediction function & Model

$$\hat{y} = \beta_0 + \beta_1 x$$

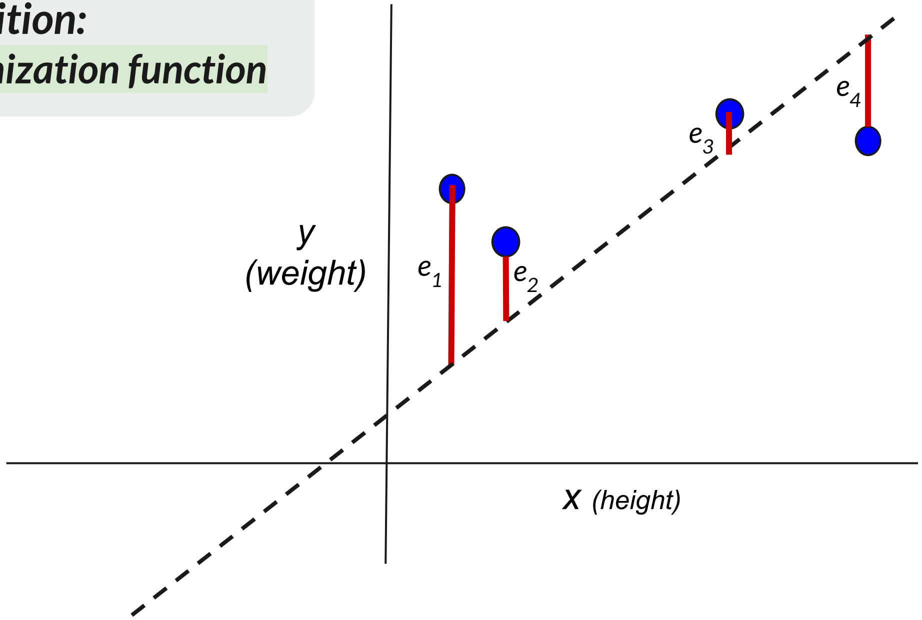
$$\Theta = \{\beta_0, \beta_1\}$$

Fitting Error

$$\boxed{\epsilon = |\hat{y} - y|}$$

Linear Regression

Definition:
Optimization function



Prediction function & Model

$$\hat{y} = \beta_0 + \beta_1 x$$

$$\Theta = \{\beta_0, \beta_1\}$$

Fitting Error

$$\epsilon = |\hat{y} - y|$$

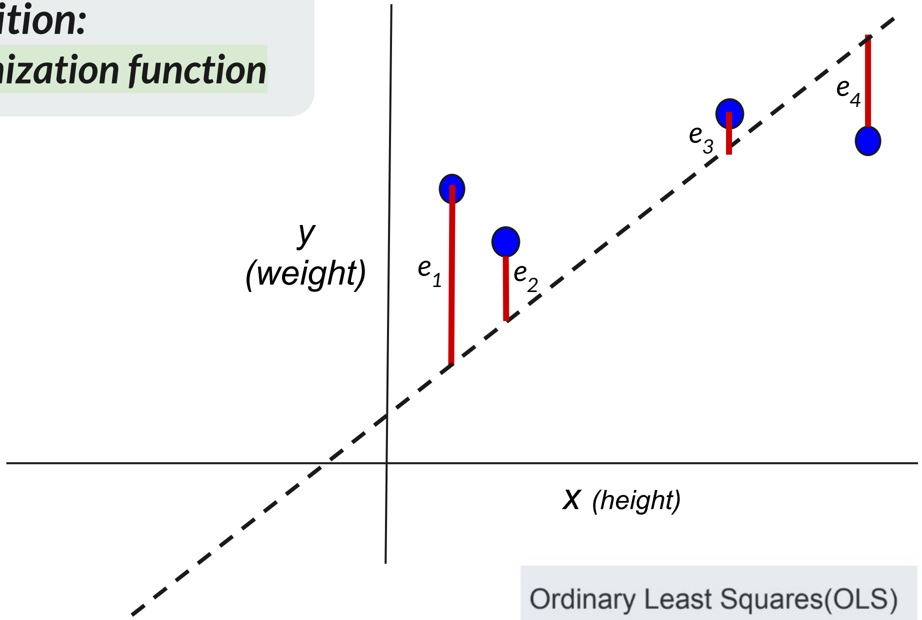
Optimization/loss/error function

$$E_\Theta = \frac{1}{2}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2$$

$$\Theta^* = \mathrm{argmin}_\Theta\, E\{(x_i, y_i)\}_{i=1,\cdots,N}$$

Linear Regression

Definition:
Optimization function



Ordinary Least Squares(OLS)

$$\hat{\beta} = (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\mathbf{y},$$

Out of scope today ….

Prediction function & Model

$$\hat{y} = \beta_0 + \beta_1 x$$

$$\Theta = \{\beta_0, \beta_1\}$$

Fitting Error

$$\epsilon = |\hat{y} - y|$$

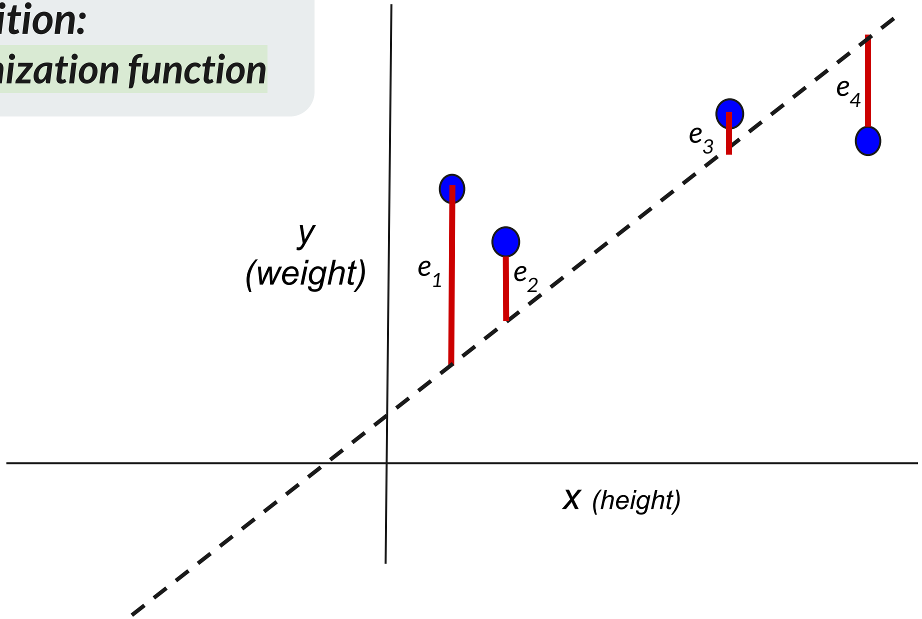
Optimization/loss/error function

$$E_\Theta = \frac{1}{2}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2$$

$$\Theta^* = \mathrm{argmin}_\Theta\, E\{(x_i, y_i)\}_{i=1,\cdots,N}$$

Linear Regression

Definition:
Optimization function

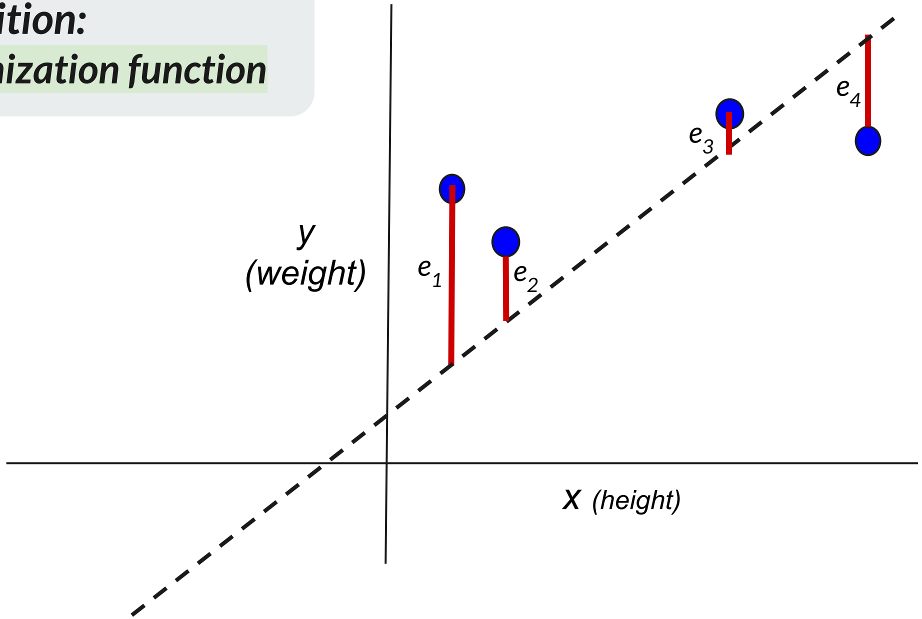


$$E_{\Theta} = \frac{1}{2}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2$$

$$E_{\Theta} = \frac{1}{2}\sum_{i=1}^{N}(\beta_0 + \beta_1 x_i - y_i)^2$$

Linear Regression

Definition:
Optimization function



- **Optimization/Loss Function:**

$$E_\Theta = \frac{1}{2}\sum_{i=1}^{N}(\beta_0 + \beta_1 x_i - y_i)^2$$

- **Gradient w.r.t.** β_0**:**

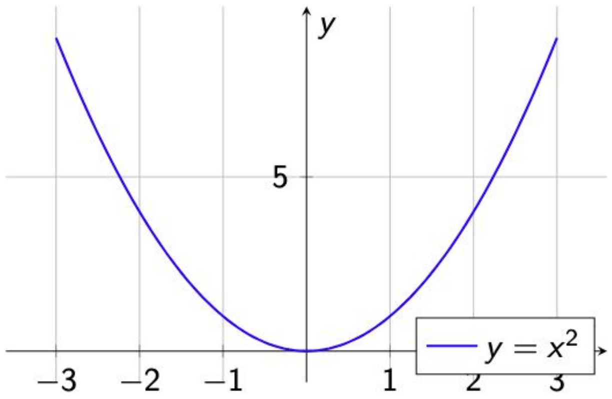
$$\frac{\partial E_\Theta}{\partial \beta_0} = \sum_{i=1}^{N}(\beta_0 + \beta_1 x_i - y_i)$$

- **Gradient w.r.t.** β_1**:**

$$\frac{\partial E_\Theta}{\partial \beta_1} = \sum_{i=1}^{N}(\beta_0 + \beta_1 x_i - y_i)\, x_i$$

Linear Regression

Definition: Function derivatives

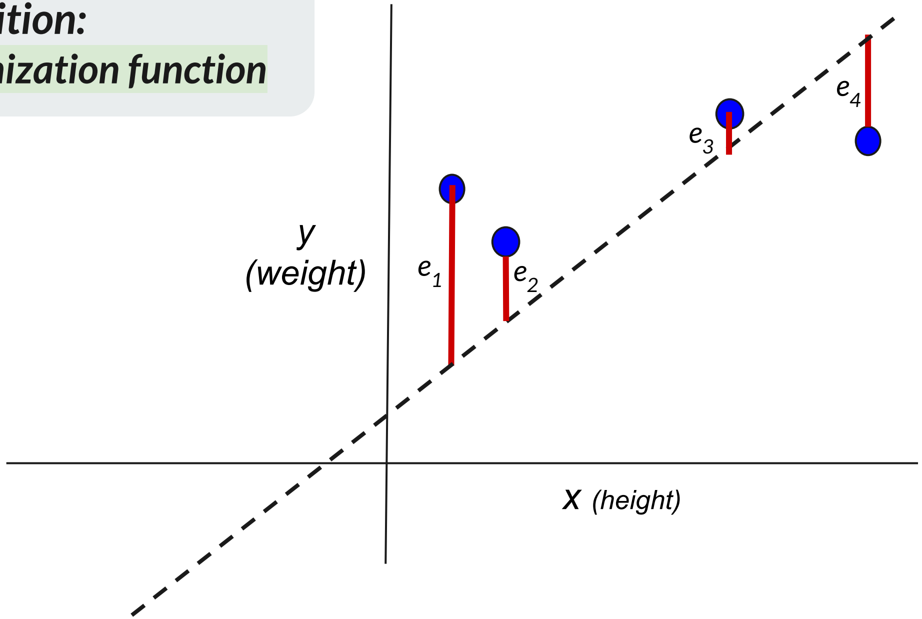


$$y = x^2 \qquad \text{(Quadratic function)}$$

$$\frac{dy}{dx} = 2x \qquad \text{(First Derivative)}$$

Linear Regression

Definition:
Optimization function



- **Optimization/Loss Function:**

$$E_\Theta = \frac{1}{2}\sum_{i=1}^{N}(\beta_0 + \beta_1 x_i - y_i)^2$$

- **Gradient w.r.t.** β_0**:**

$$\frac{\partial E_\Theta}{\partial \beta_0} = \sum_{i=1}^{N}(\beta_0 + \beta_1 x_i - y_i)$$

- **Gradient w.r.t.** β_1**:**

$$\frac{\partial E_\Theta}{\partial \beta_1} = \sum_{i=1}^{N}(\beta_0 + \beta_1 x_i - y_i)\, x_i$$

QA