COMP2261 ARTIFICIAL INTELLIGENCE / MACHINE LEARNING

Linear Regression
-- Cost Function

Dr SHI Lei

Learning Objectives

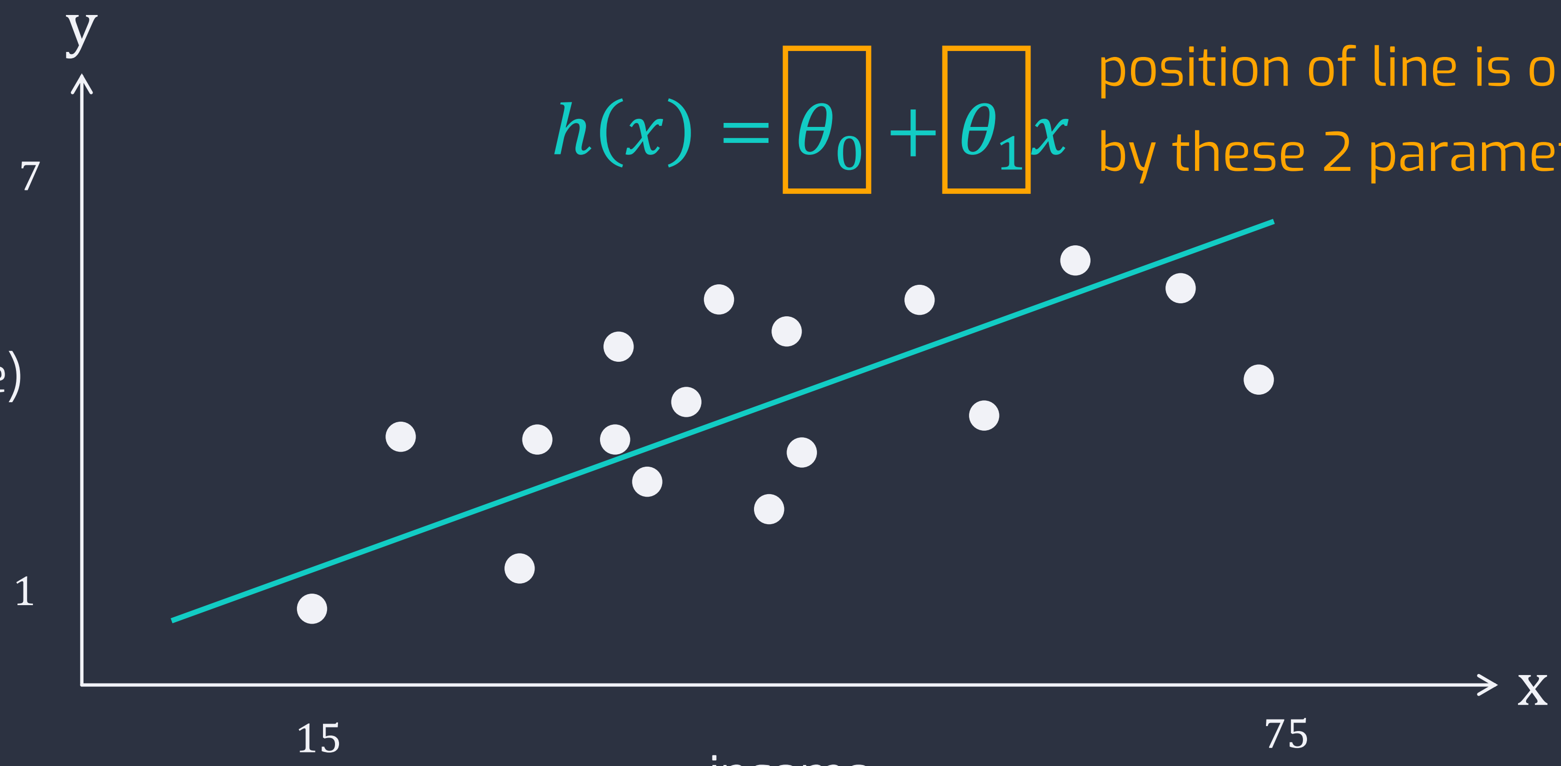
- Understand the notion of cost
- Understand what is a cost function
- Understand the relationship between a hypothesis function and its corresponding cost function

EXAMPLE. annual income to predict happiness

$$h(x) = \boxed{\theta_0} + \boxed{\theta_1} x$$

position of line is only determined by these 2 parameters

happiness
(dependent variable)

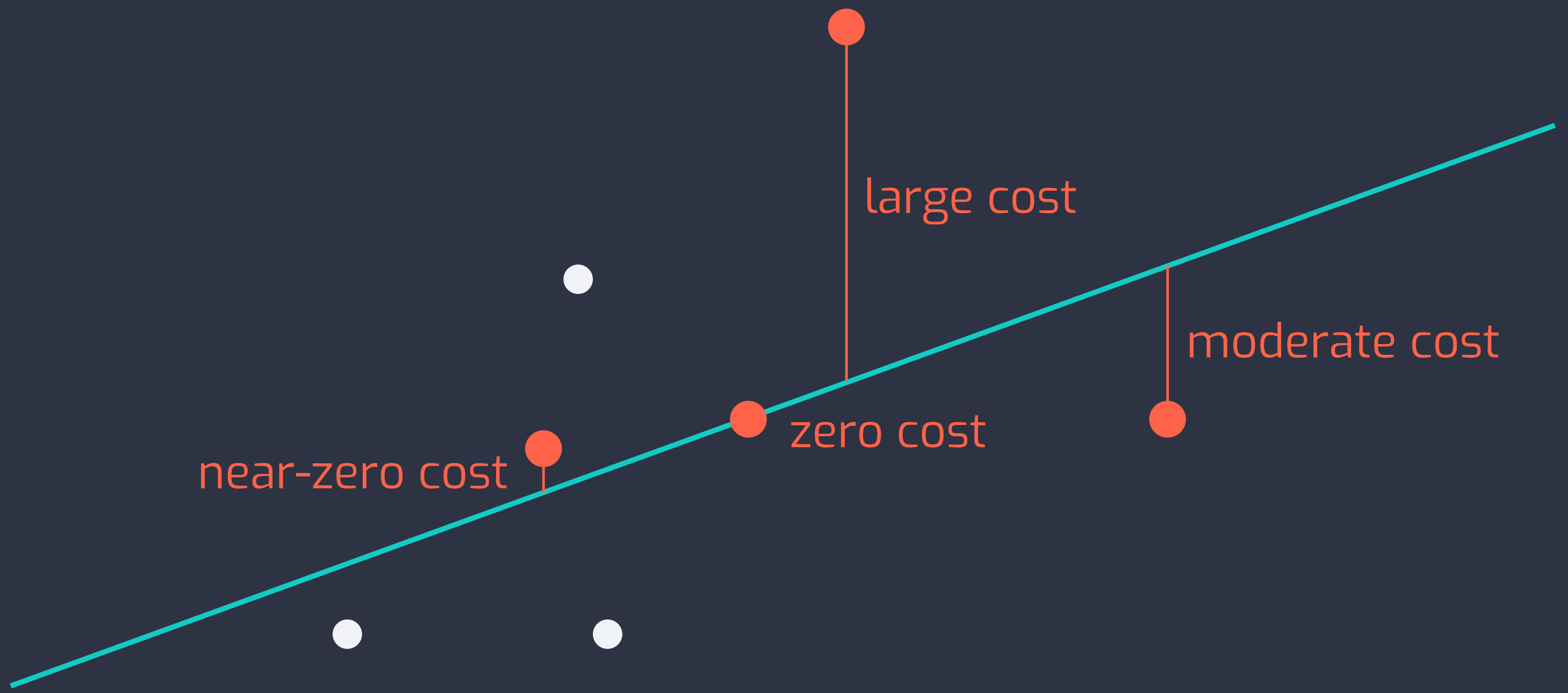


income
(independent variable)

Cost

Cost

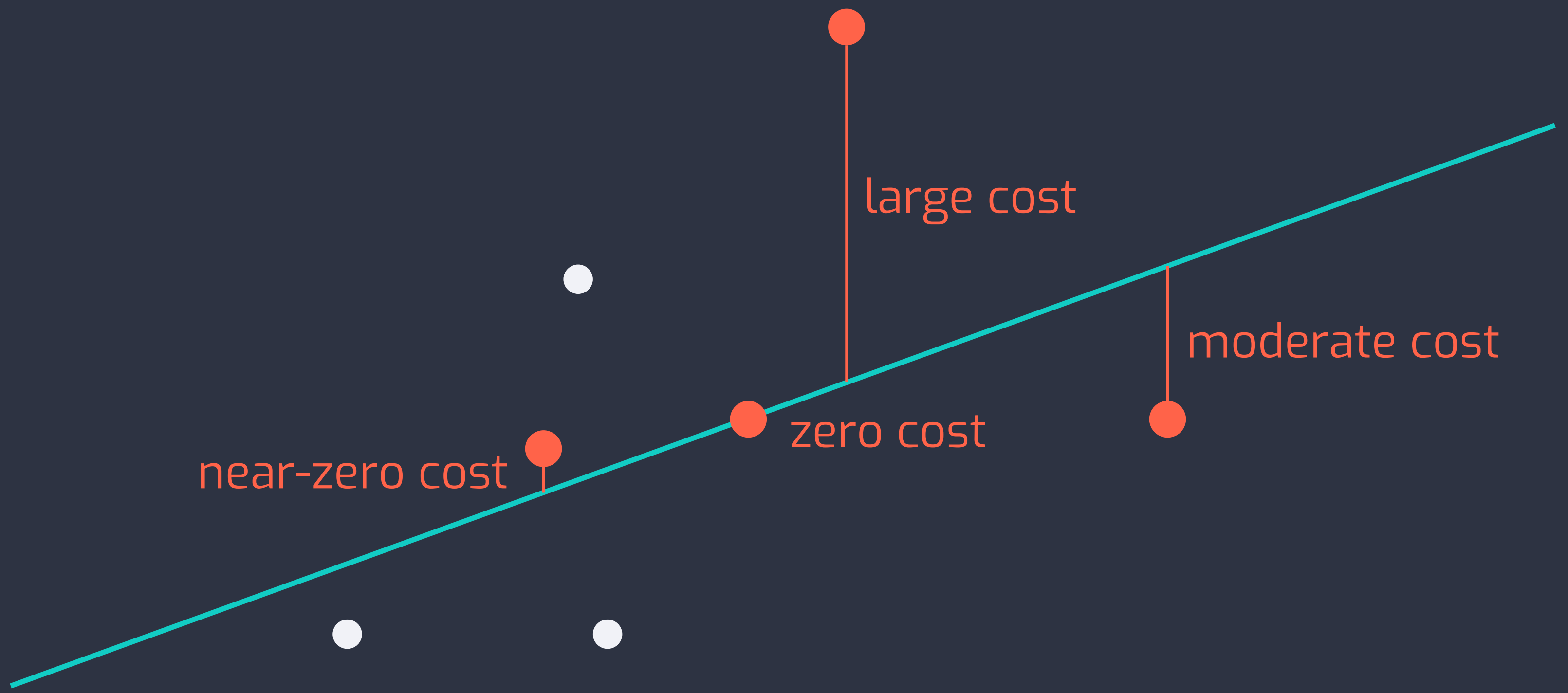
- Measure of difference between prediction and actual value for a given x.
- Distance from point on model (predicted) to point from the training set (actual).
- Value is either zero or positive.
- We want straight line so that not only one or a few but all having small/zero cost.



Cost

- We want straight line so that not only one or a few but all having small/zero cost.
- Squared Error Cost Function

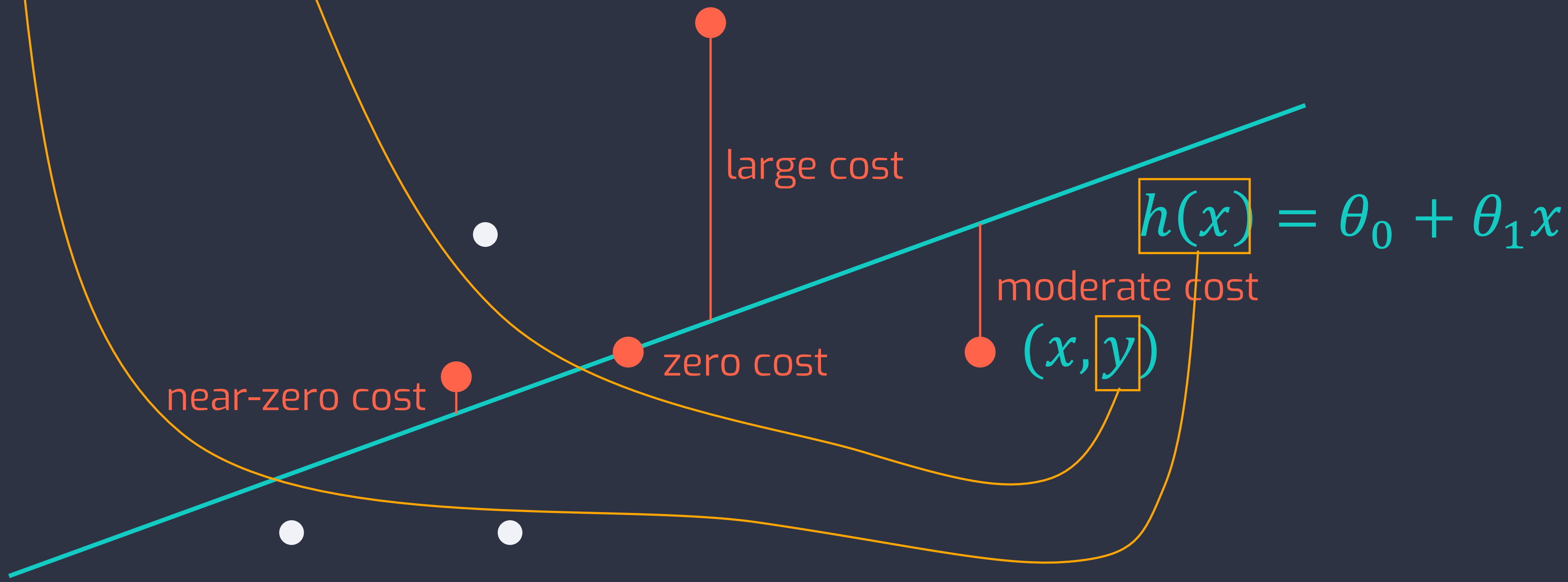
$$\boxed{\sum_{i=1}^{m}}(predicted^{(i)} - actual^{(i)})^2$$



Cost

- We want straight line so that not only one or a few but all having small/zero cost.
- Squared Error Cost Function

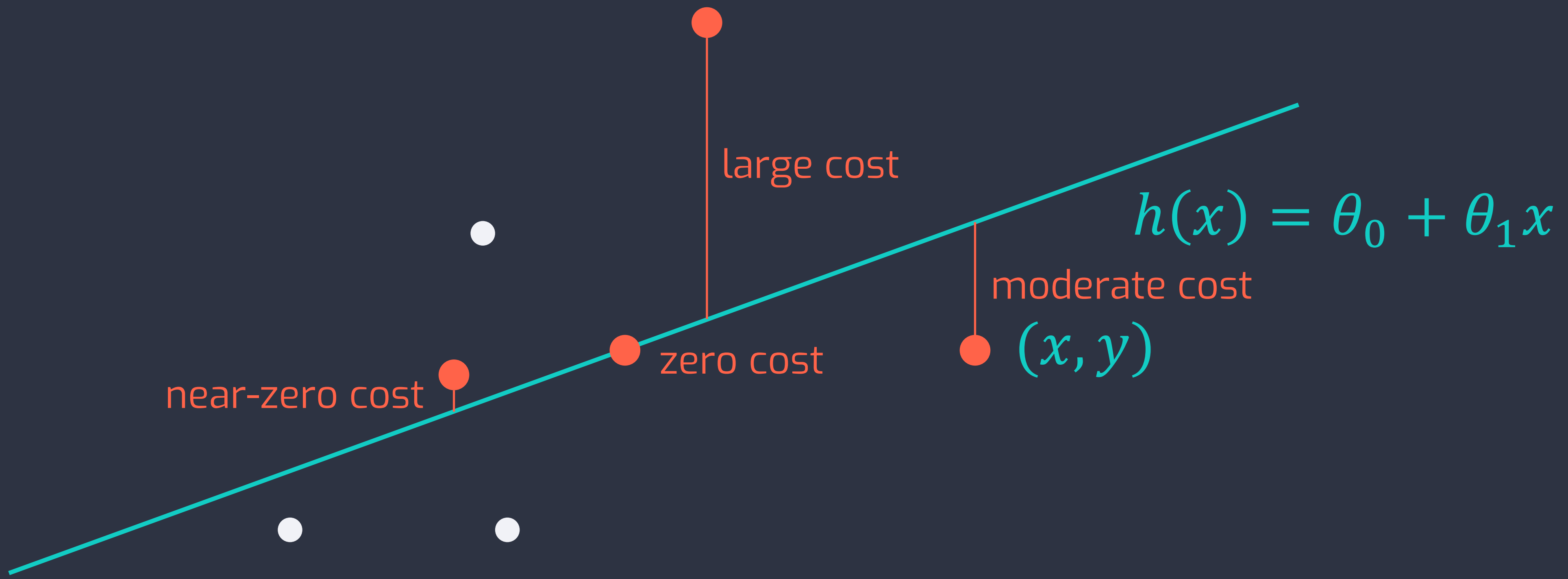
$$\sum_{i=1}^{m}\left(predicted^{(i)} - actual^{(i)}\right)^2 = \sum_{i=1}^{m}\left(h(x)^{(i)} - y^{(i)}\right)^2$$



Cost

- We want straight line so that not only one or a few but all having small/zero cost.
- Squared Error Cost Function

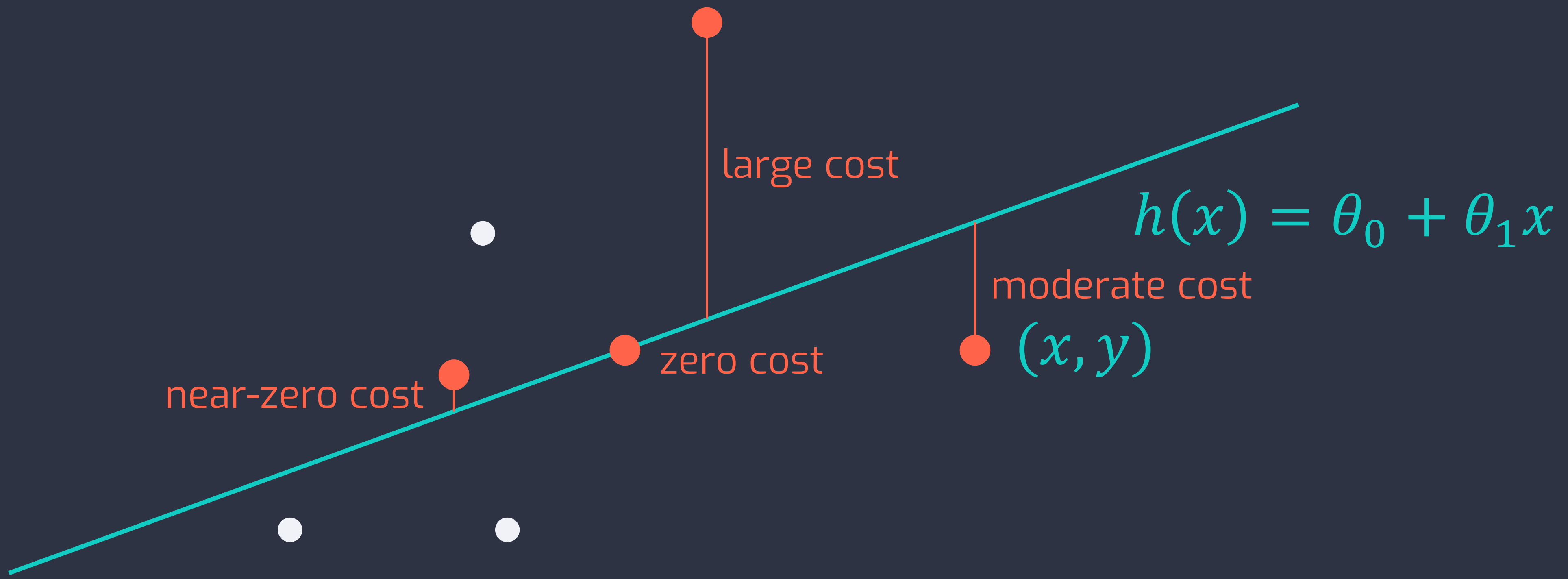
$$\sum_{i=1}^{m}(predicted^{(i)} - actual^{(i)})^2 = \boxed{\sum_{i=1}^{m}}(h(x)^{(i)} - y^{(i)})^2$$



Cost

- We want straight line so that not only one or a few but all having small/zero cost.
- Squared Error Cost Function

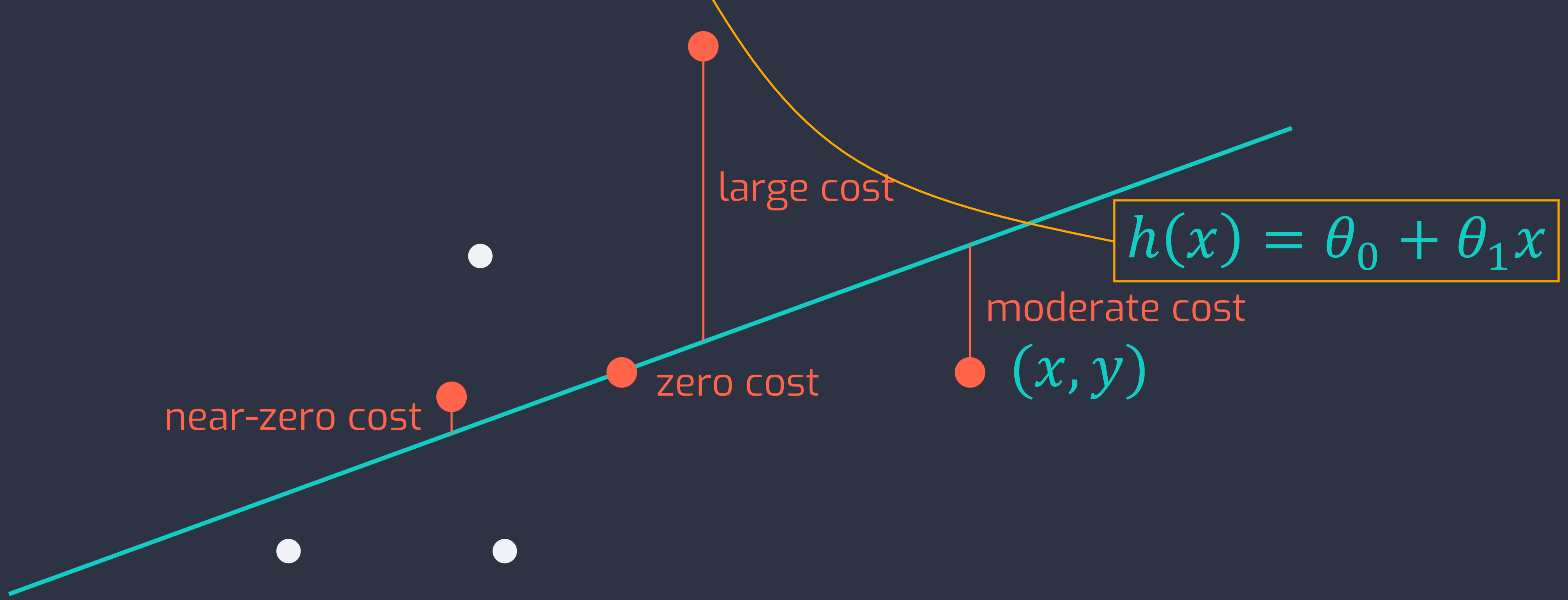
$$\sum_{i=1}^{m}(predicted^{(i)} - actual^{(i)})^2 = \boxed{\frac{1}{2m}}\sum_{i=1}^{m}(h(x)^{(i)} - y^{(i)})^2$$



Cost

- We want straight line so that not only one or a few but all having small/zero cost.
- Squared Error Cost Function

$$\sum_{i=1}^{m}(predicted^{(i)} - actual^{(i)})^2 = \frac{1}{2m}\sum_{i=1}^{m}(\boxed{h(x)}^{(i)} - y^{(i)})^2 = \frac{1}{2m}\sum_{i=1}^{m}(\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2$$



Cost

- We want straight line so that not only one or a few but all having small/zero cost.
- Squared Error Cost Function

Goal: $$\underset{\theta_0,\theta_1}{\text{minimise}} = \frac{1}{2m}\sum_{i=1}^{m}\left(\boxed{\theta_0} + \boxed{\theta_1}x^{(i)} - y^{(i)}\right)^2$$

$$h(x) = \boxed{\theta_0} + \boxed{\theta_1}x$$

Cost

- We want straight line so that not only one or a few but all having small/zero cost.
- Squared Error Cost Function

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2$$

$$h(x) = \theta_0 + \theta_1 x$$

Hypothesis Function

$$h(x) = \theta_0 + \theta_1 x$$

- Represents the <u>model</u>
- Values of θ_0 and θ_1 are fixed
- x is the independent variable
- A function of independent variable x

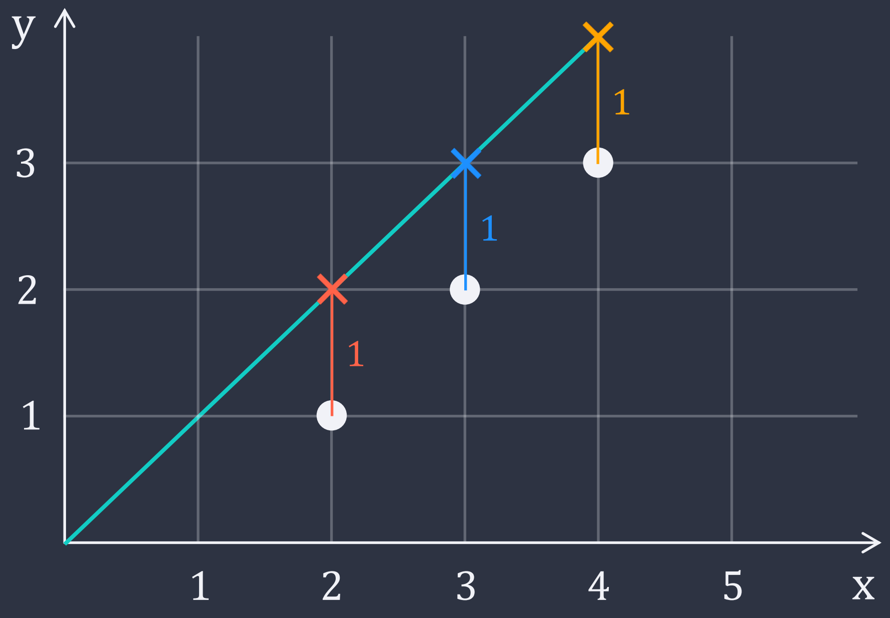
Cost Function

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2$$

- Represents the <u>cost</u>
- Values of $x^{(i)}$ and $y^{(i)}$ are fixed
- θ_0 and θ_1 are the independent variables
- A function of independent variables θ_1, θ_2

Hypothesis Function

$$h(x) = \theta_0 + \theta_1 x$$

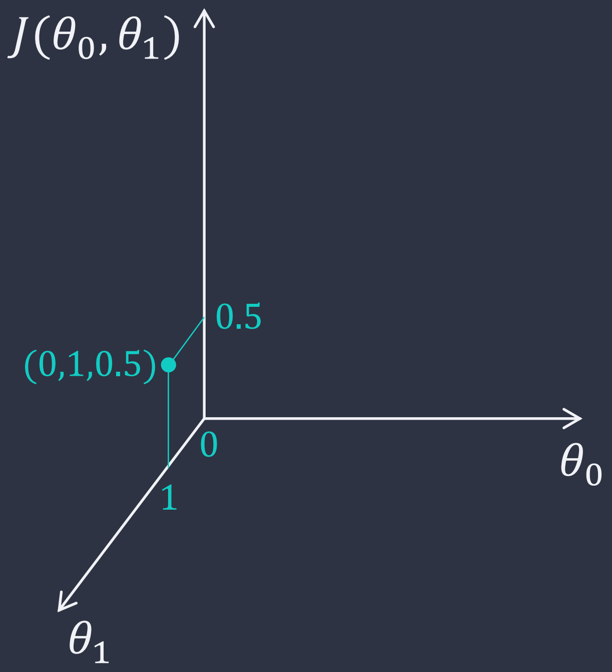


$$if\ (\theta_0 = 0, \theta_1 = 1)$$

$$J(0,1) = \frac{1}{2\times3}(1^2 + 1^2 + 1^2) = 0.5$$

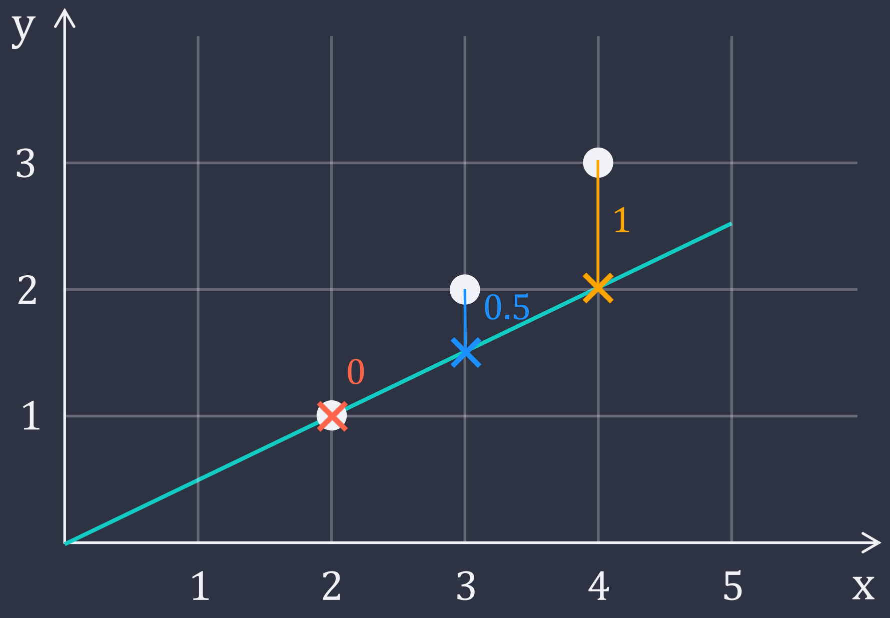
Cost Function

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2$$



Hypothesis Function

$$h(x) = \theta_0 + \theta_1 x$$

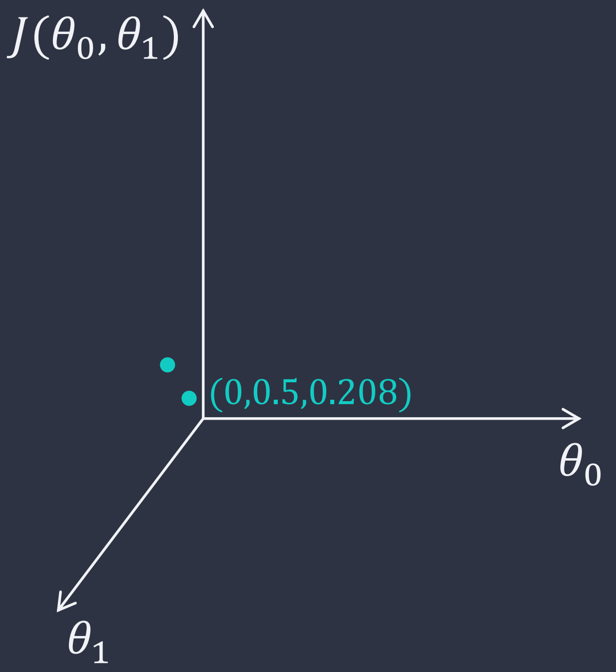


$$if\ (\theta_0 = 0, \theta_1 = 0.5)$$

$$J(0,1) = \frac{1}{2\times3}(0^2 + 0.5^2 + 1^2) = 0.208$$

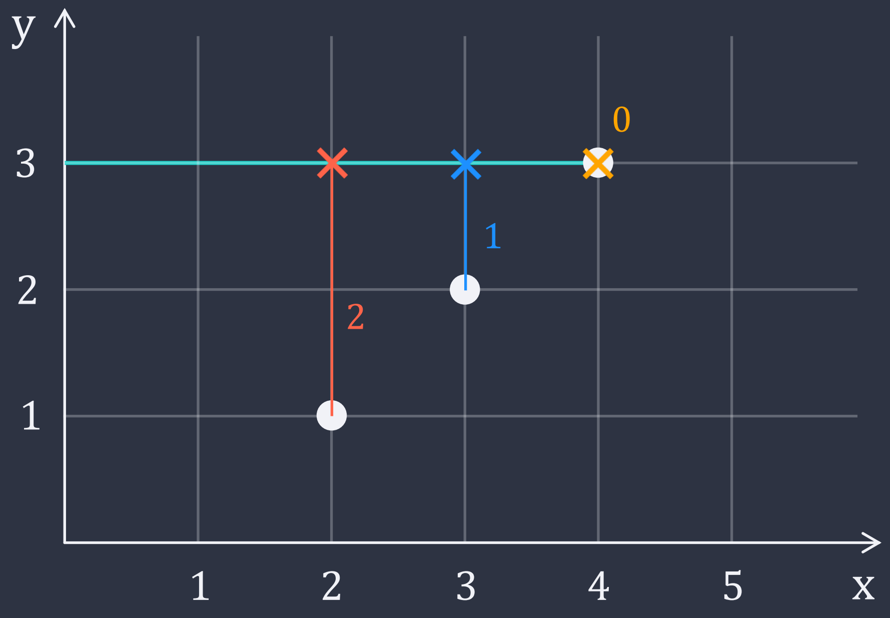
Cost Function

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2$$



Hypothesis Function

$$h(x) = \theta_0 + \theta_1 x$$

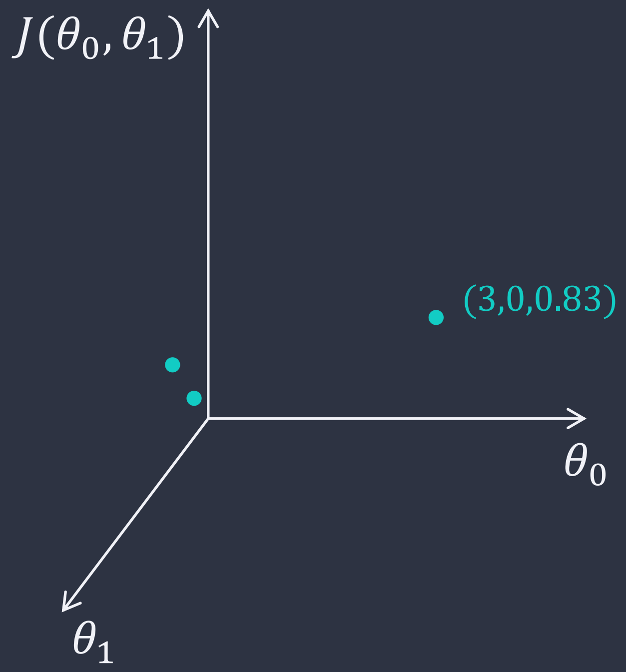


$$if\ (\theta_0 = 3, \theta_1 = 0)$$

$$J(0,1) = \frac{1}{2\times 3}(2^2 + 1^2 + 0^2) = 0.83$$

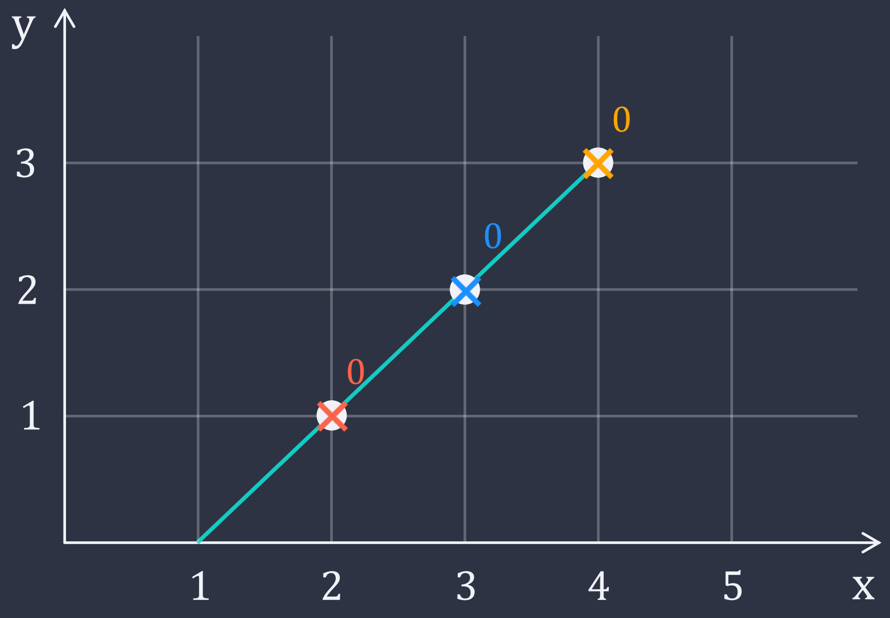
Cost Function

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2$$



Hypothesis Function

$$h(x) = \theta_0 + \theta_1 x$$

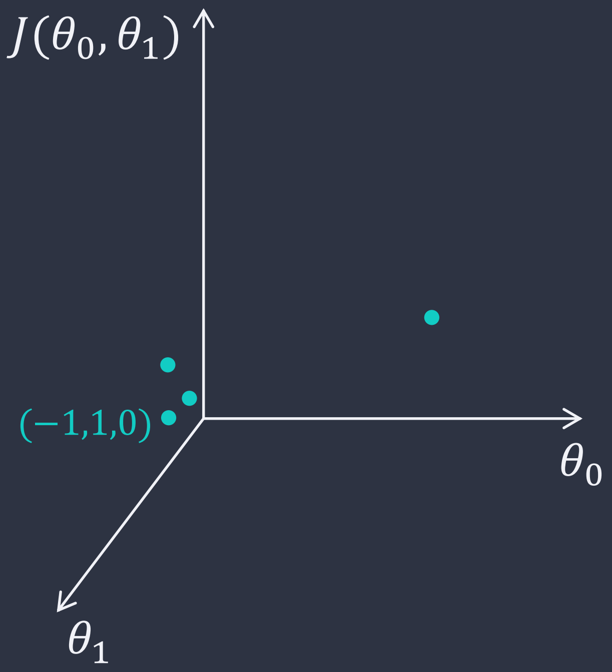


$$if\ (\theta_0 = -1, \theta_1 = 1)$$

$$J(0,1) = \frac{1}{2\times3}(0^2 + 0^2 + 0^2) = 0$$

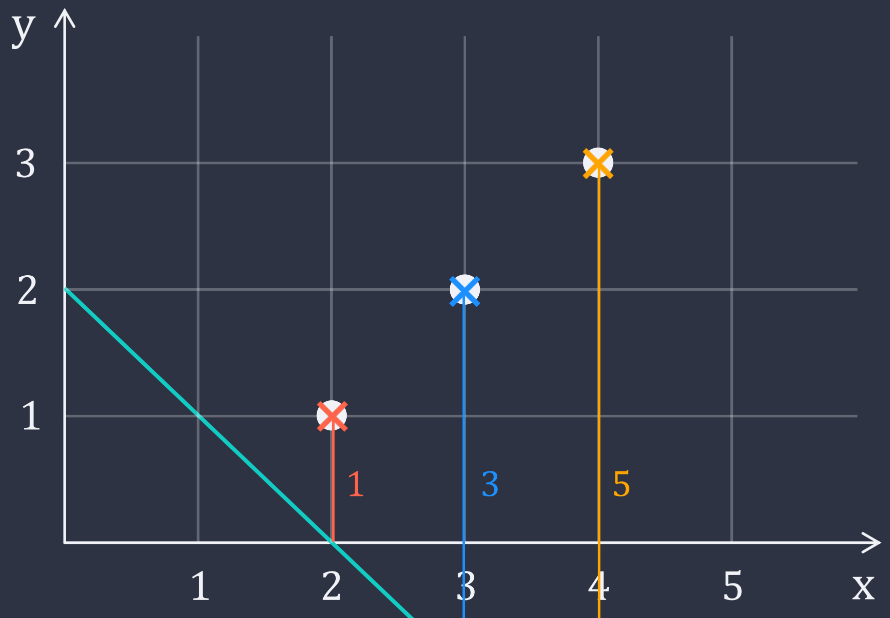
Cost Function

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2$$



Hypothesis Function

$$h(x) = \theta_0 + \theta_1 x$$

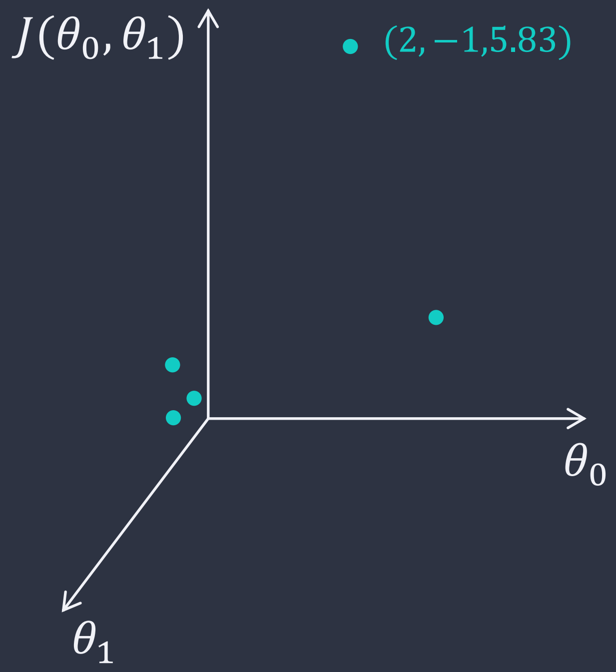


$$if\ (\theta_0 = 2, \theta_1 = -1)$$

$$J(0,1) = \frac{1}{2\times3}(1^2 + 3^2 + 5^2) = 5.83$$

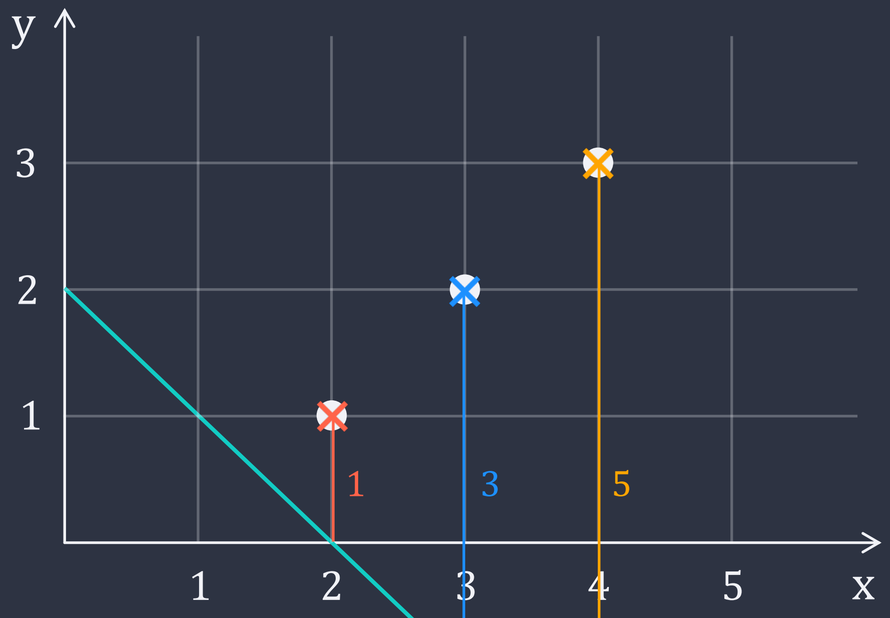
Cost Function

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2$$



Hypothesis Function

$$h(x) = \theta_0 + \theta_1 x$$

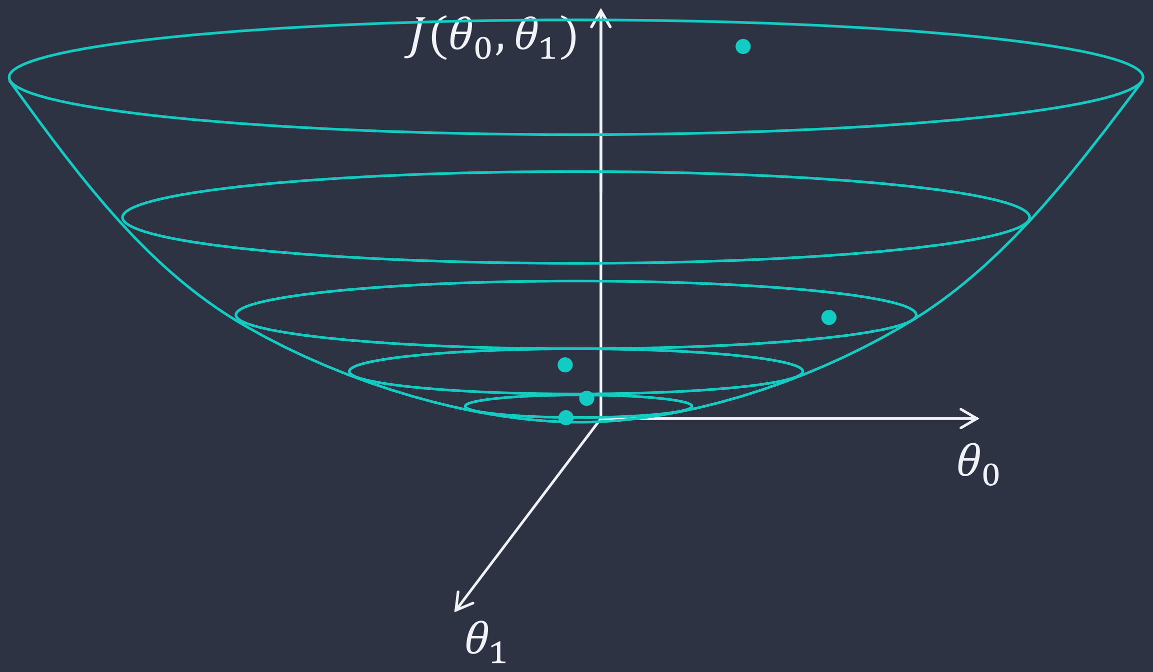


$$if\ (\theta_0 = 2, \theta_1 = -1)$$

$$J(0,1) = \frac{1}{2 \times 3}(1^2 + 3^2 + 5^2) = 5.83$$

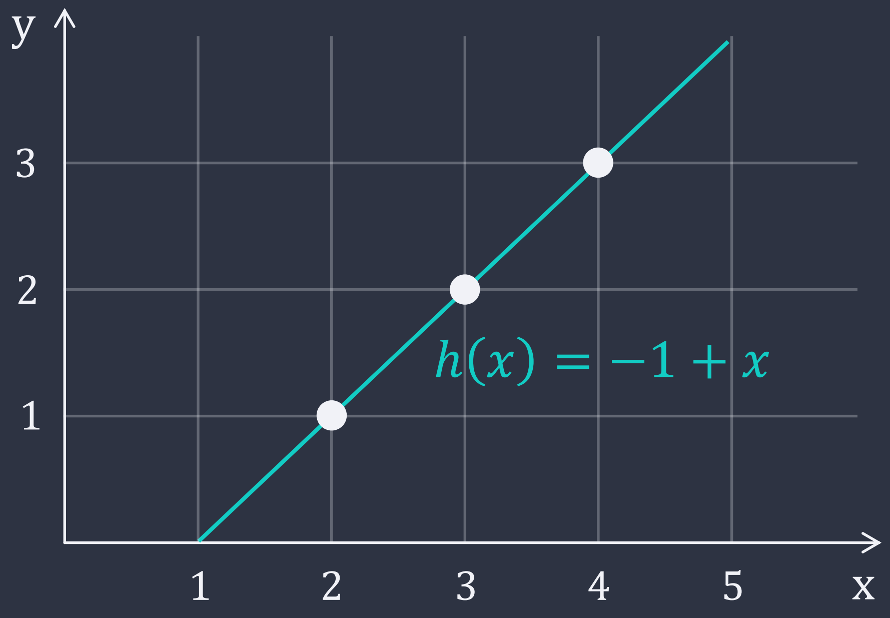
Cost Function

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2$$



Hypothesis Function

$$h(x) = \theta_0 + \theta_1 x$$

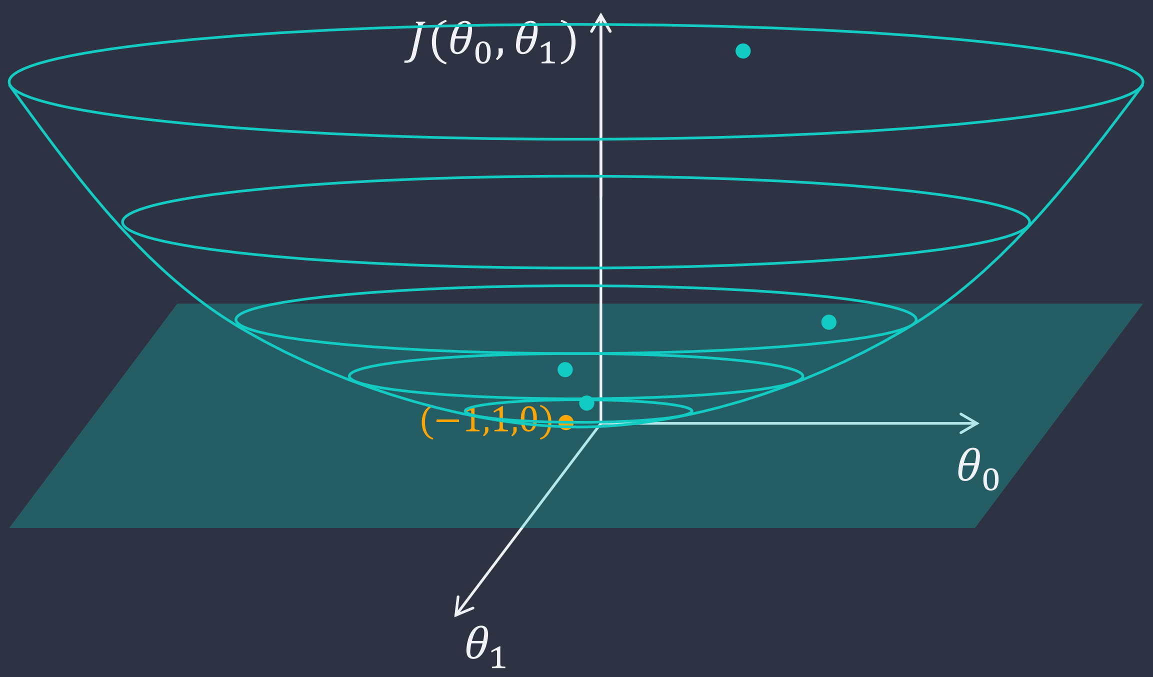


$\theta_0 = -1, \theta_1 = 1$

$J(0,1) = 0$

Cost Function

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2$$



Takeaway Points

- Cost is the measure of the difference between the predicted value and the actual value.
- The goal of a linear regression algorithm is to find the parameter pair so that the cost is minimised.
- We use the cost function $J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2$ to help find the best possible parameter pair.