SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

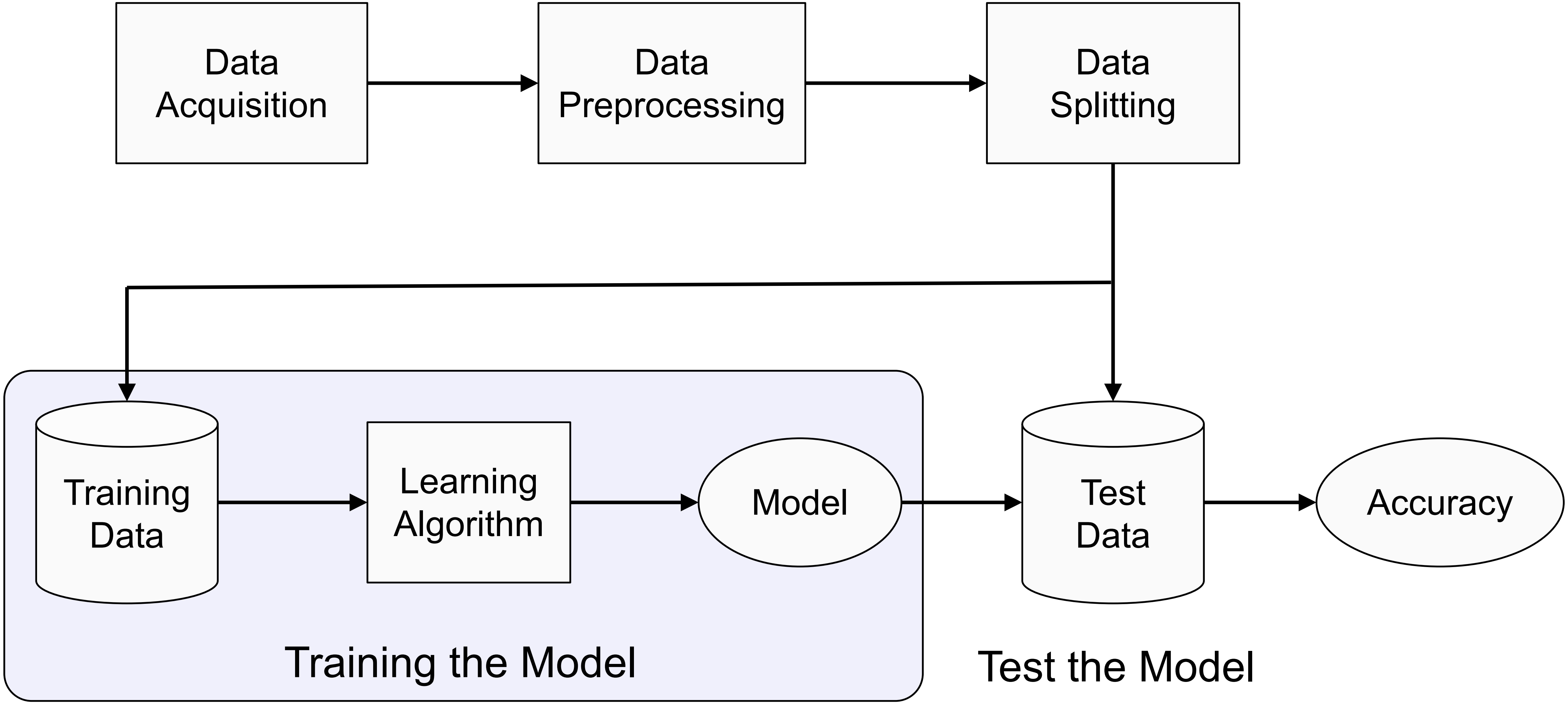
10.020 Data Driven World

Linear Regression: Testing

Peng Song, ISTD

Week 9, Lesson 2, 2021

Revision: Linear Regression



Revision: Linear Regression

- Hypothesis: $h_\beta(x) = \beta_0 + \beta_1 x$
- Cost function: $J(\beta_0, \beta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\beta\left(x^{(i)}\right) - y^{(i)}\right)^2$
- Gradient Descent:

$$\beta_0 \coloneqq \beta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\beta\left(x^{(i)}\right) - y^{(i)}\right)$$

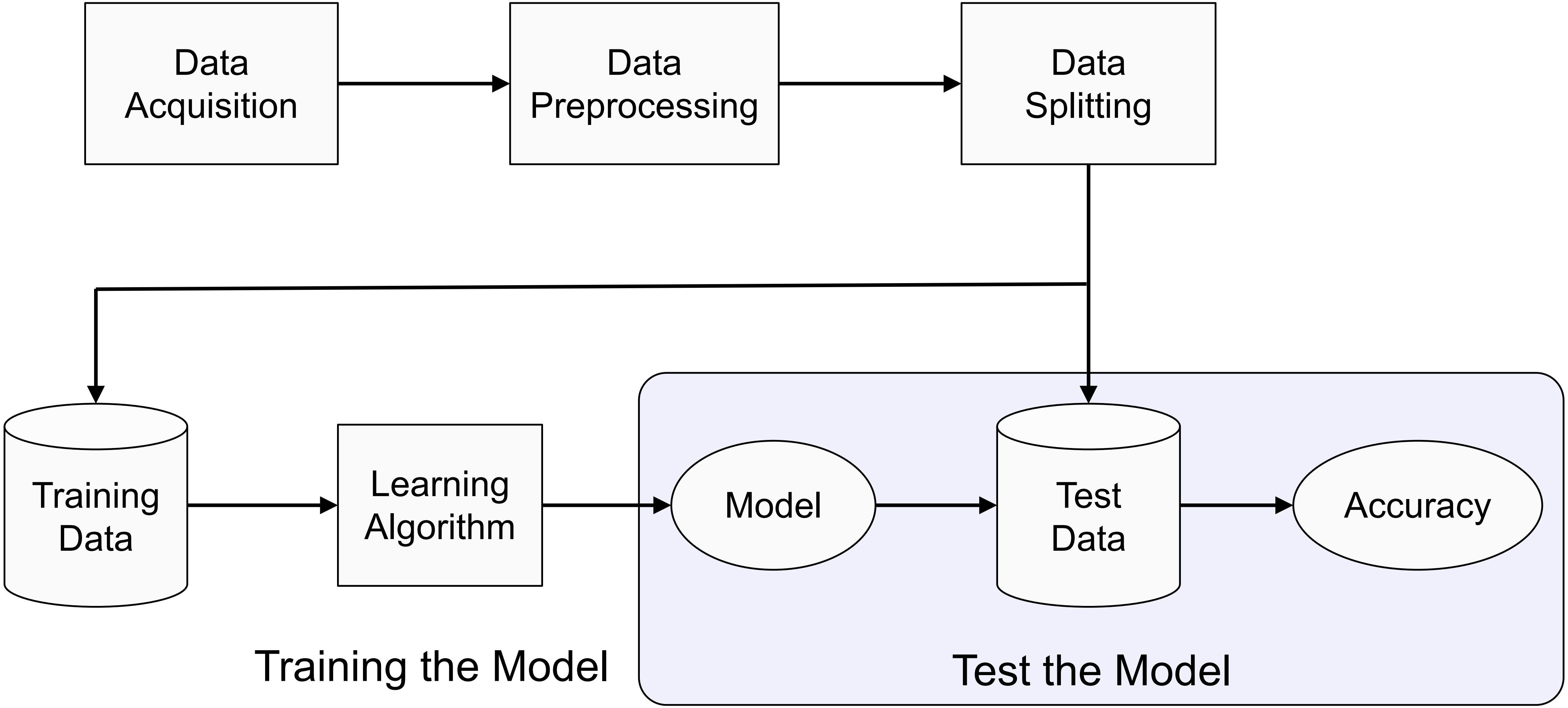
$$\beta_1 \coloneqq \beta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\beta\left(x^{(i)}\right) - y^{(i)}\right) x^{(i)}$$

Revision: Linear Regression

- Hypothesis: $\hat{\boldsymbol{y}} = \boldsymbol{X} \times \boldsymbol{b}$
- Cost Function: $J(\beta_0, \beta_1) = \frac{1}{2m}(\hat{\boldsymbol{y}} - \boldsymbol{y})^{\mathrm{T}} \times (\hat{\boldsymbol{y}} - \boldsymbol{y})$
- Gradient Descent: $\boldsymbol{b} = \boldsymbol{b} - \alpha \frac{1}{m} \boldsymbol{X}^{\mathrm{T}} \times (\boldsymbol{X} \times \boldsymbol{b} - \boldsymbol{y})$

where $\hat{\boldsymbol{y}} = \begin{bmatrix} \hat{y}^1 \\ \hat{y}^2 \\ \hat{y}^3 \\ .. \\ \hat{y}^m \end{bmatrix}$ $\quad \boldsymbol{X} = \begin{bmatrix} 1 & x^1 \\ 1 & x^2 \\ .. & .. \\ 1 & x^m \end{bmatrix}$ $\quad \boldsymbol{b} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ $\quad \boldsymbol{y} = \begin{bmatrix} y^1 \\ y^2 \\ y^3 \\ .. \\ y^m \end{bmatrix}$

Linear Regression: Testing



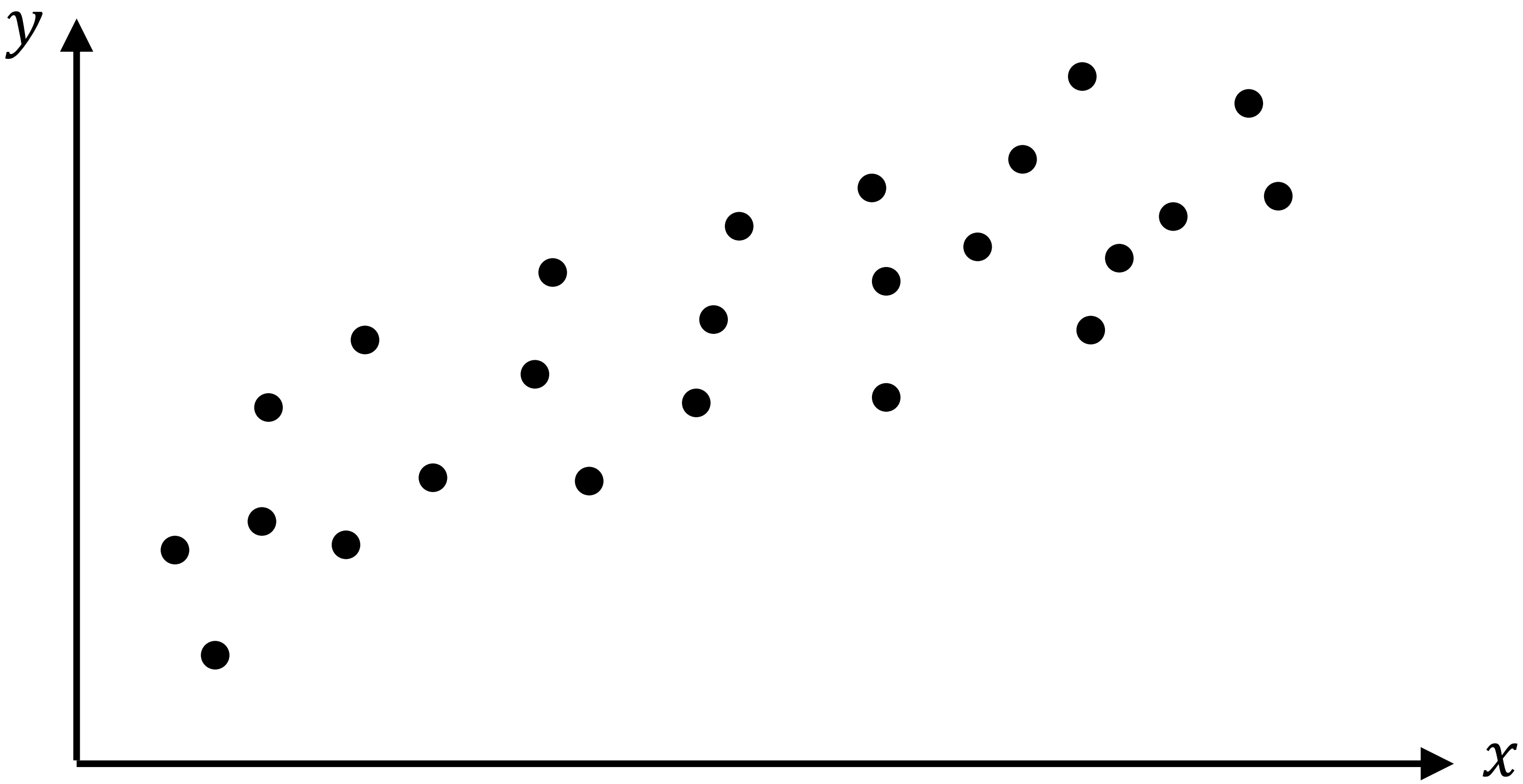
House Pricing Prediction

- Data Acquisition and Preprocessing

Size in feet^2 (x)	Price ($) in 1000’s (y)
2104	460
1416	232
1534	315
852	178
…	…

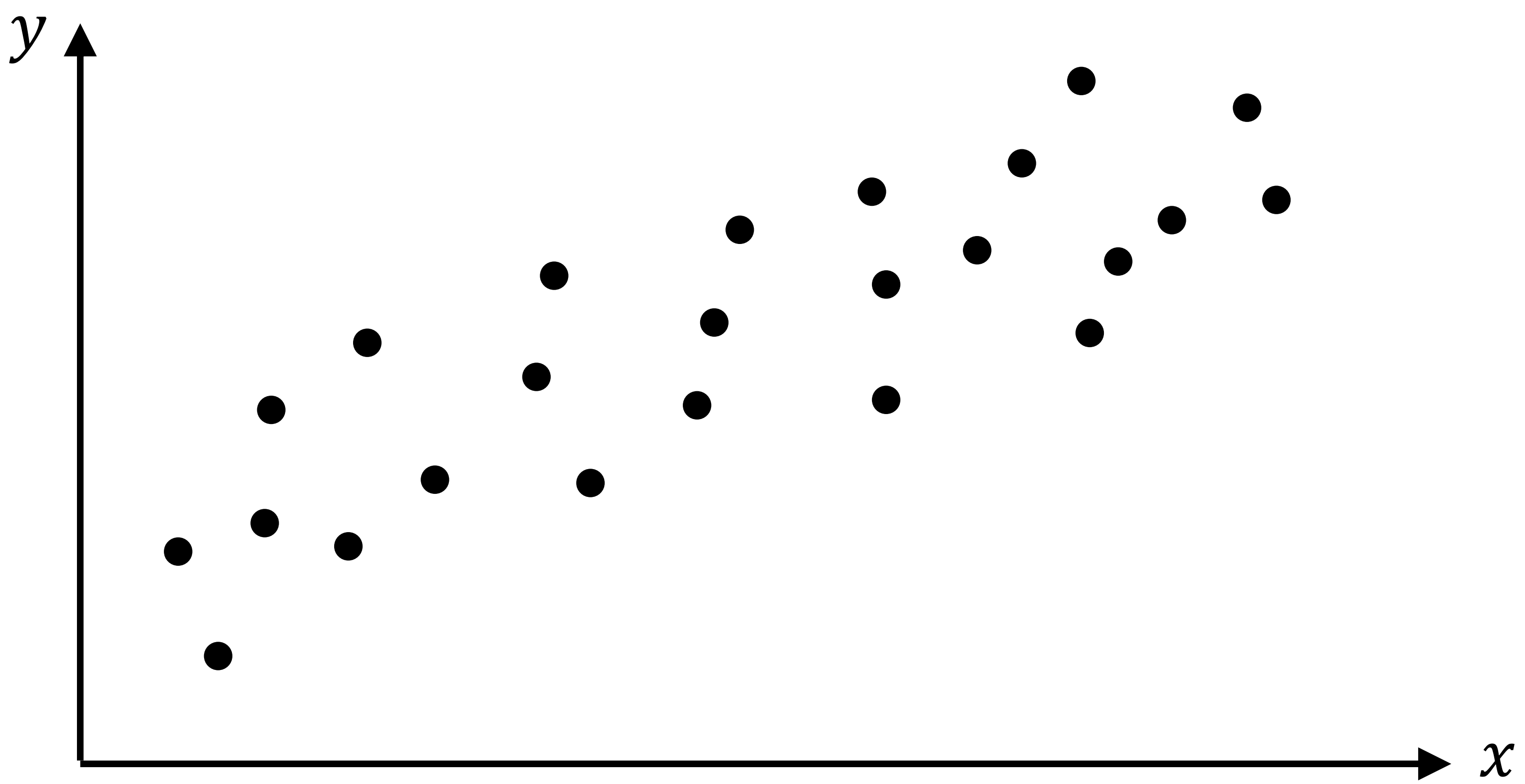
House Pricing Prediction

- Visualize preprocessed data as a scatter plot



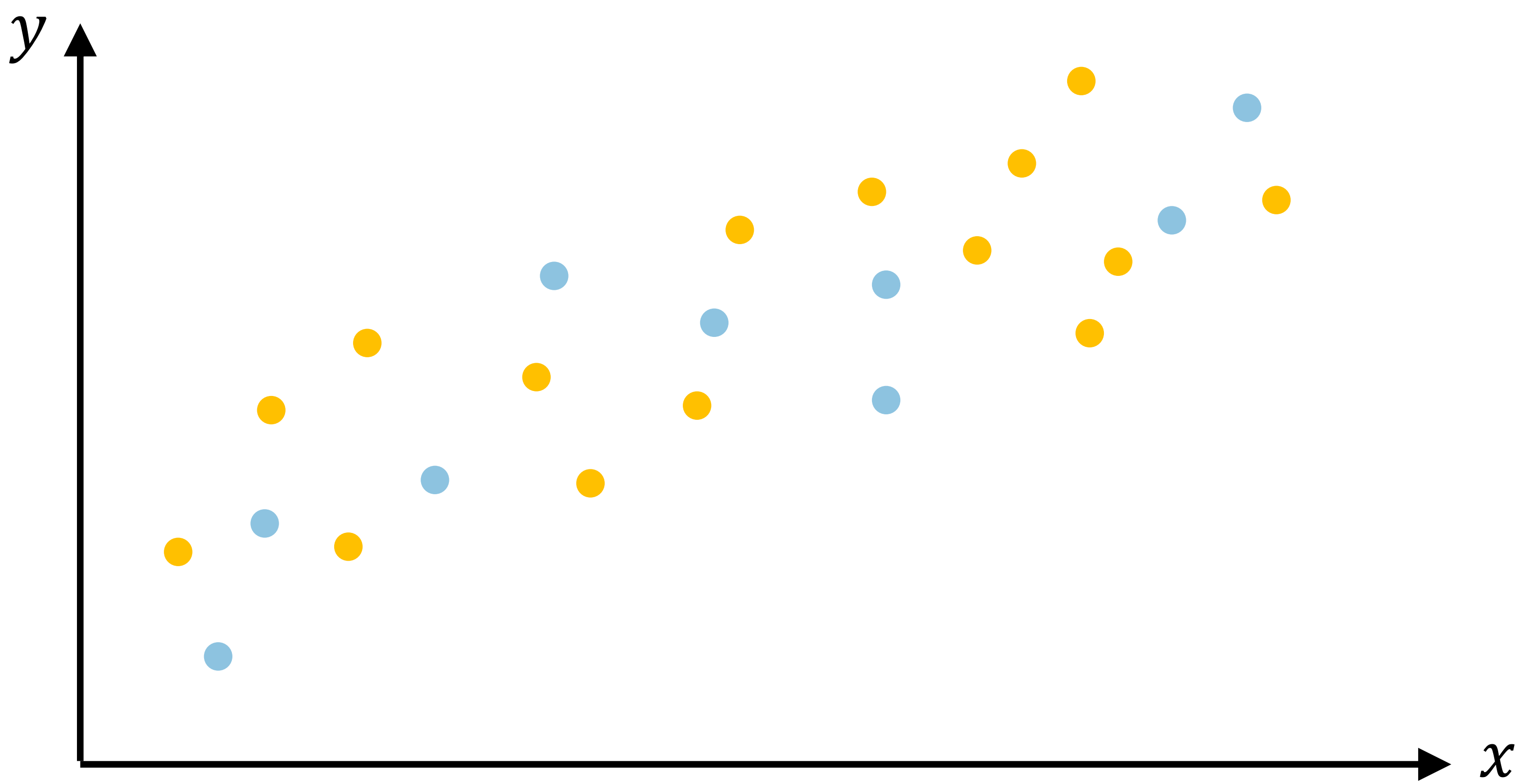
Data Splitting

- Split data into a training dataset and a testing dataset randomly



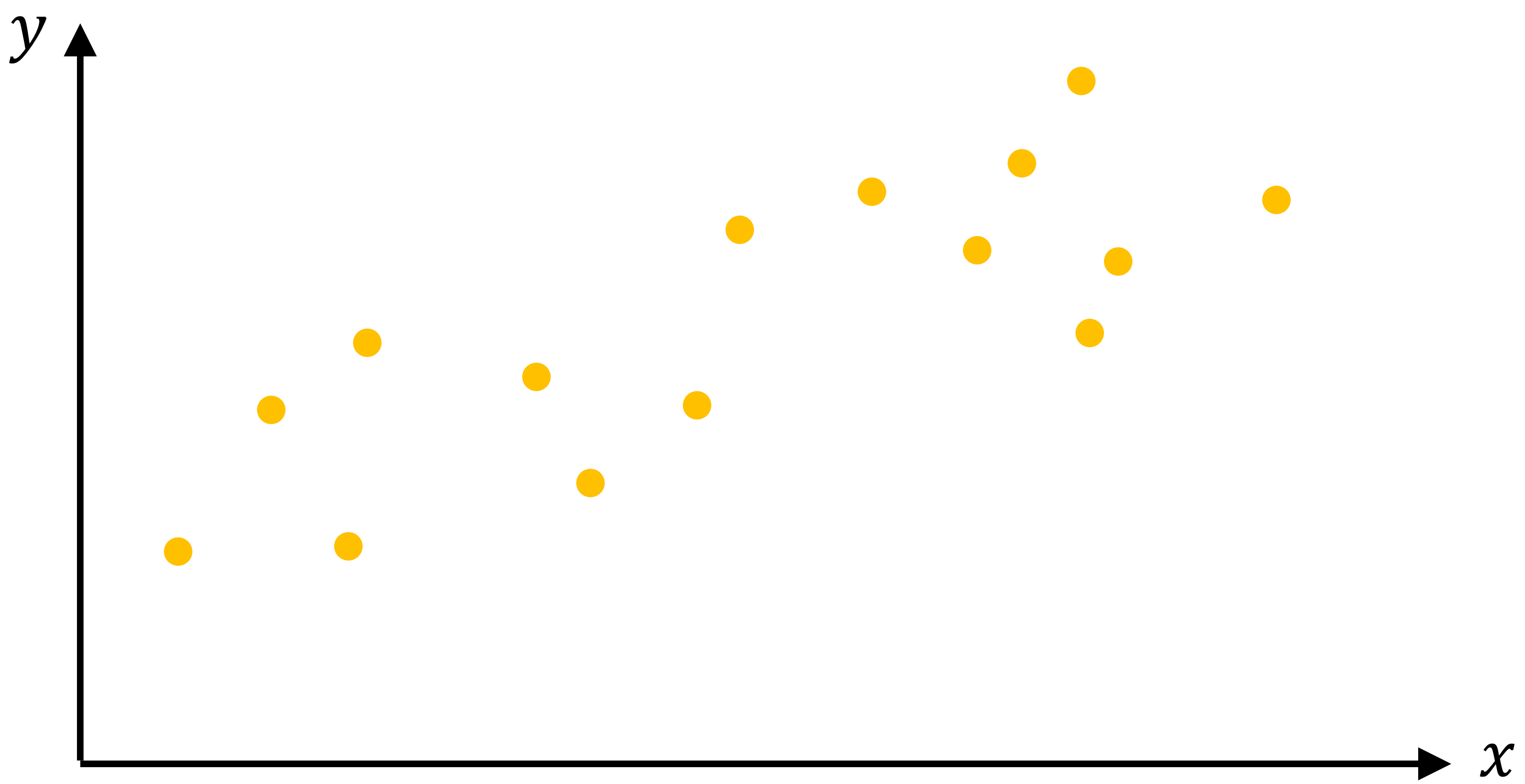
Data Splitting

- Split data into a training dataset and a testing dataset randomly



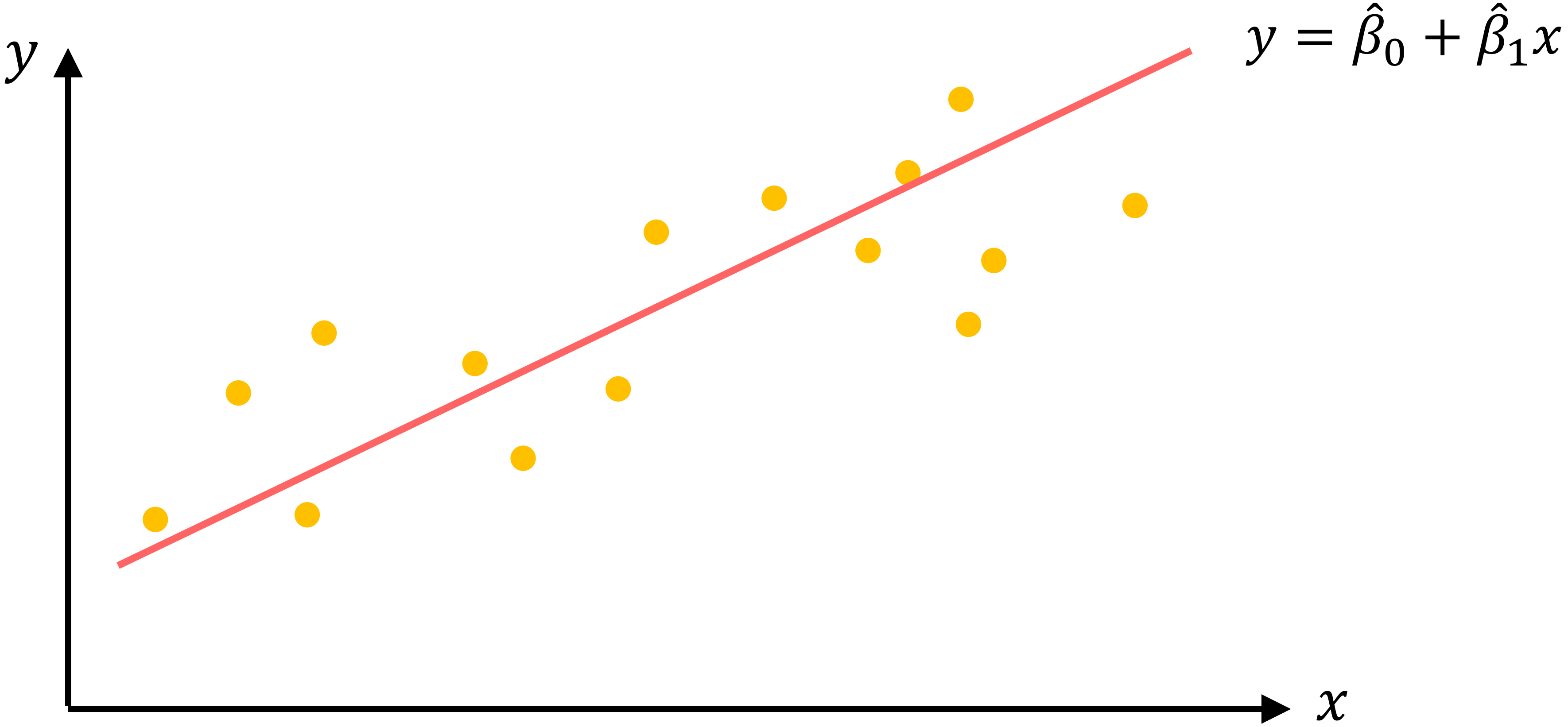
Linear Regression: Training

- Hypothesis: $h_\beta(x) = \beta_0 + \beta_1 x$



Linear Regression: Training

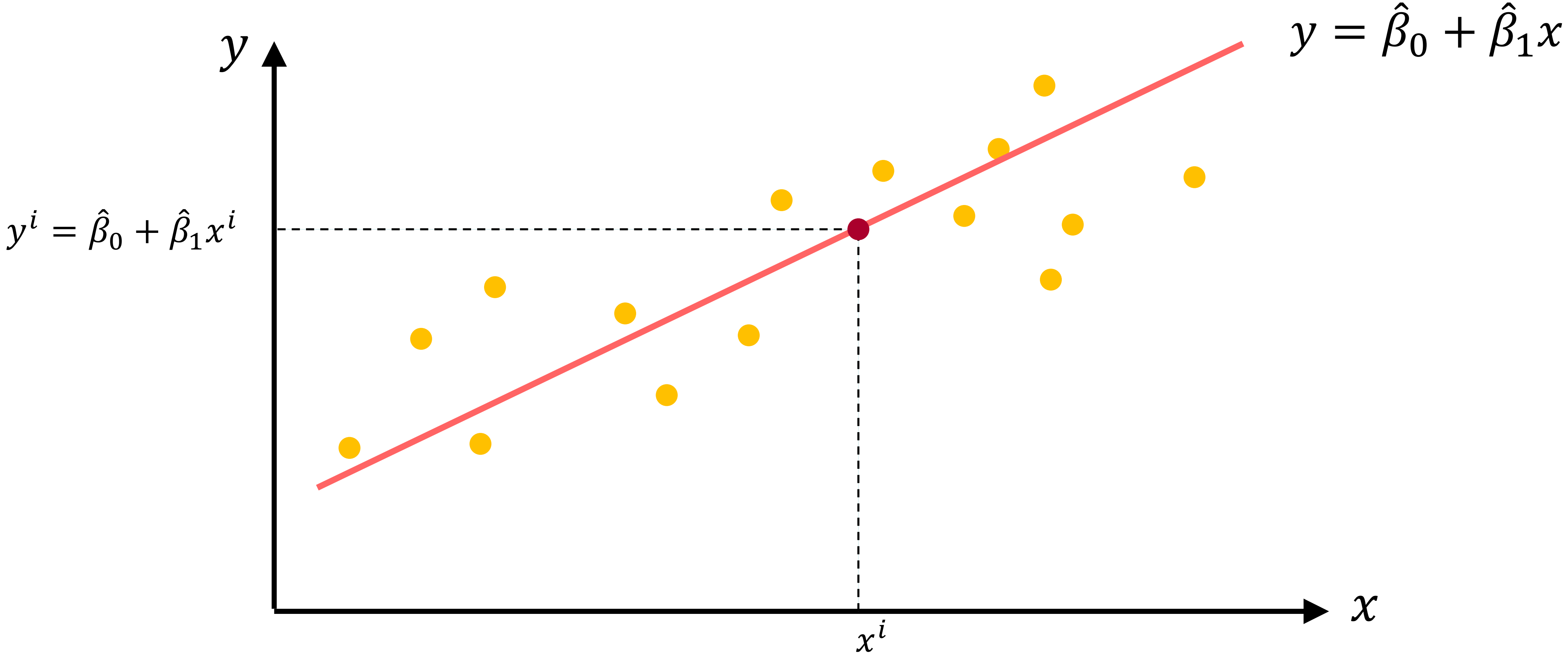
- Hypothesis: $h_\beta(x) = \beta_0 + \beta_1 x$
- Determine β_0, β_1 in the hypothesis using a gradient descent method



Linear Regression: Prediction

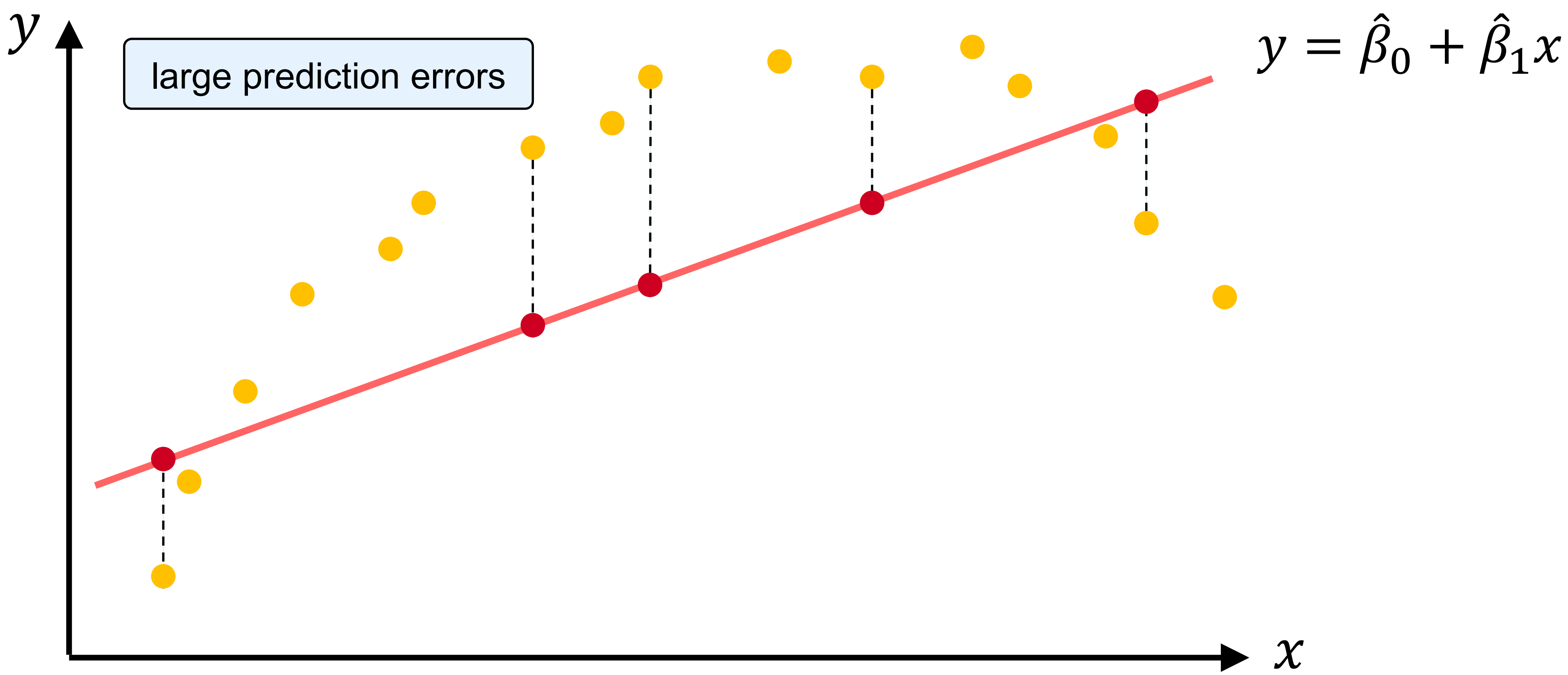
- Predict y according to a given x using the model $y = \hat{\beta}_0 + \hat{\beta}_1 x$

How to assess whether our prediction is good or not?



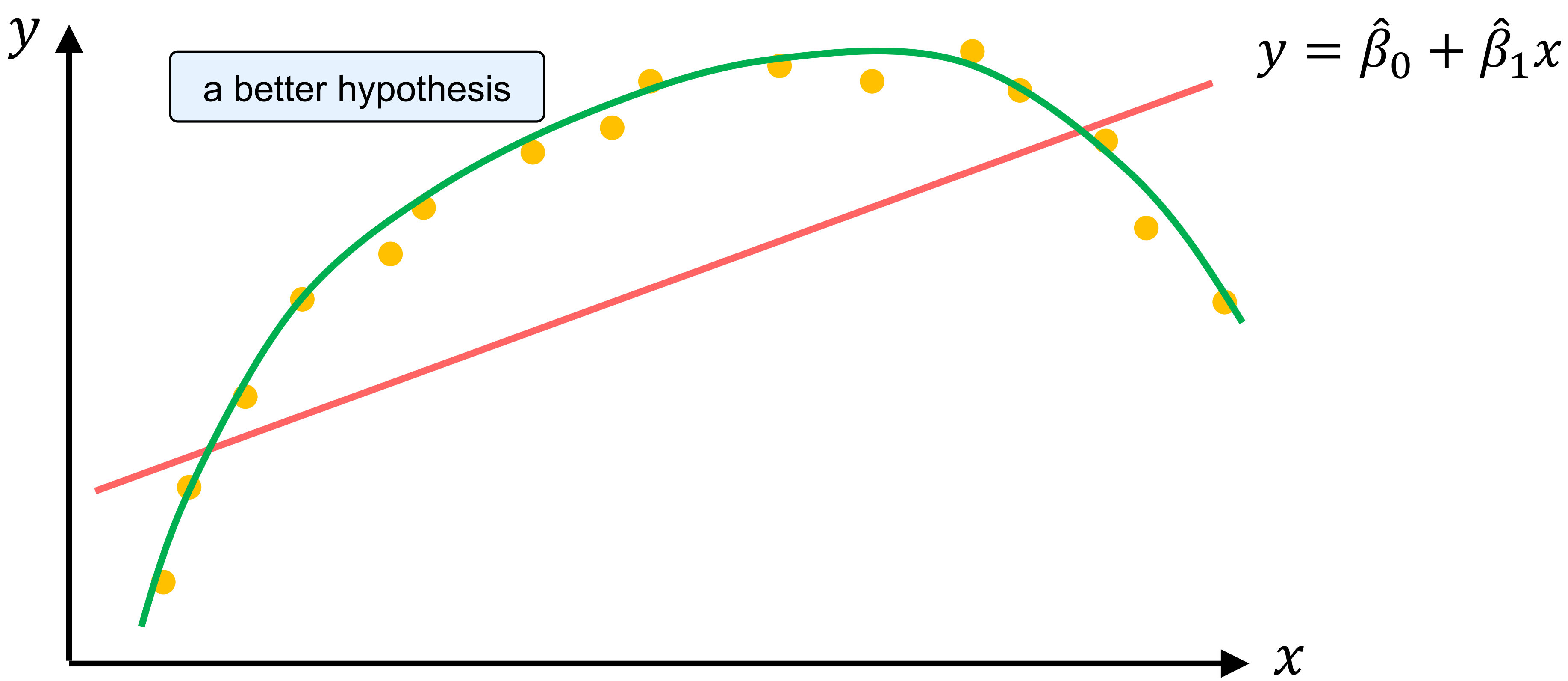
Learned Model May Not be Useful

- Case 1: relation between y and x is not linear



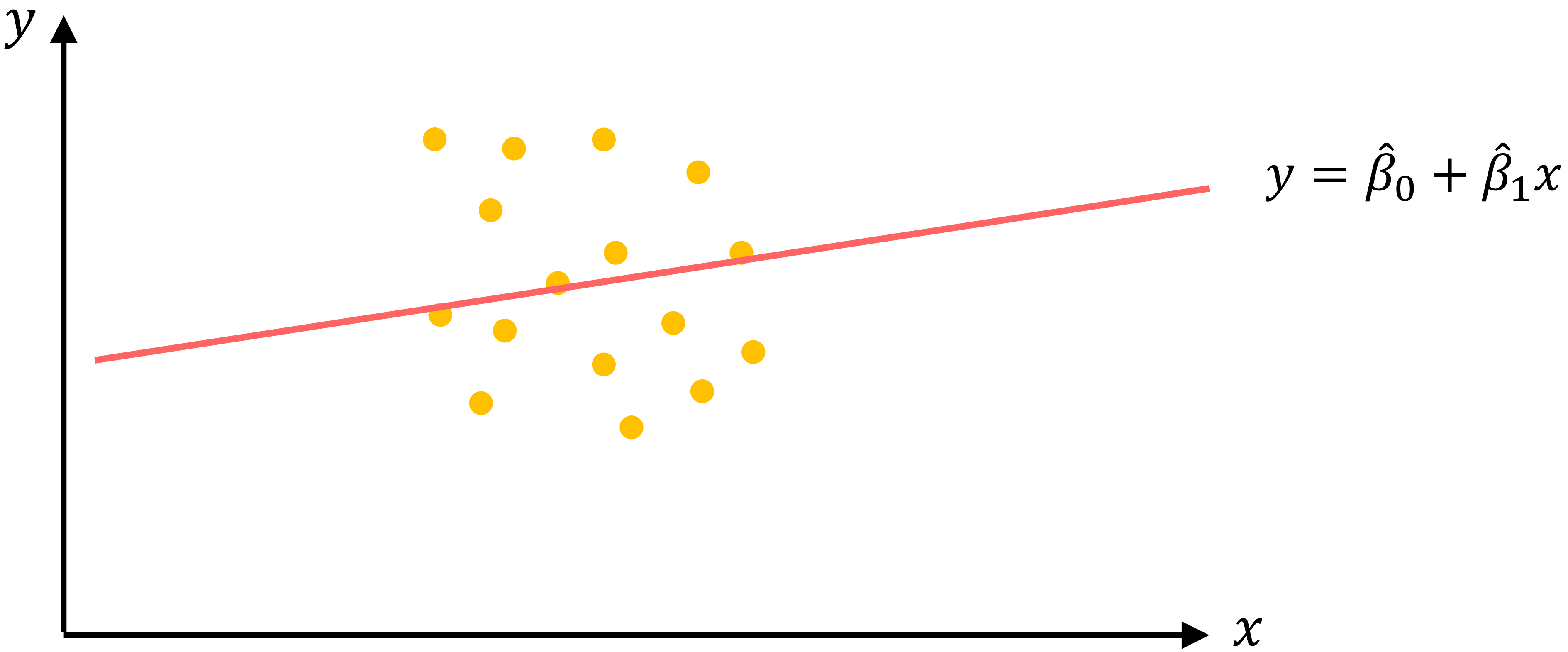
Learned Model May Not be Useful

- Case 1: relation between y and x is not linear



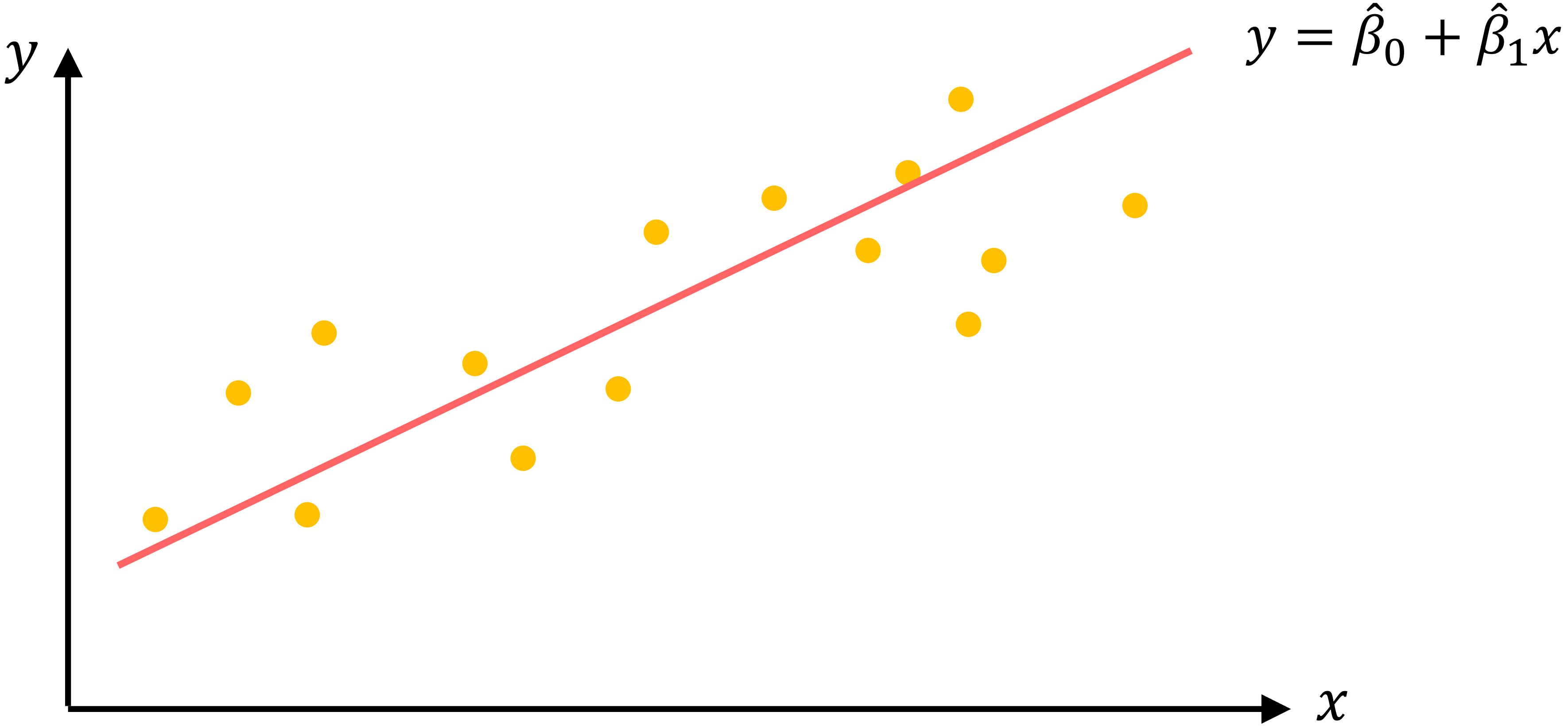
Learned Model May Not be Useful

- Case 2: no correlation between y and x



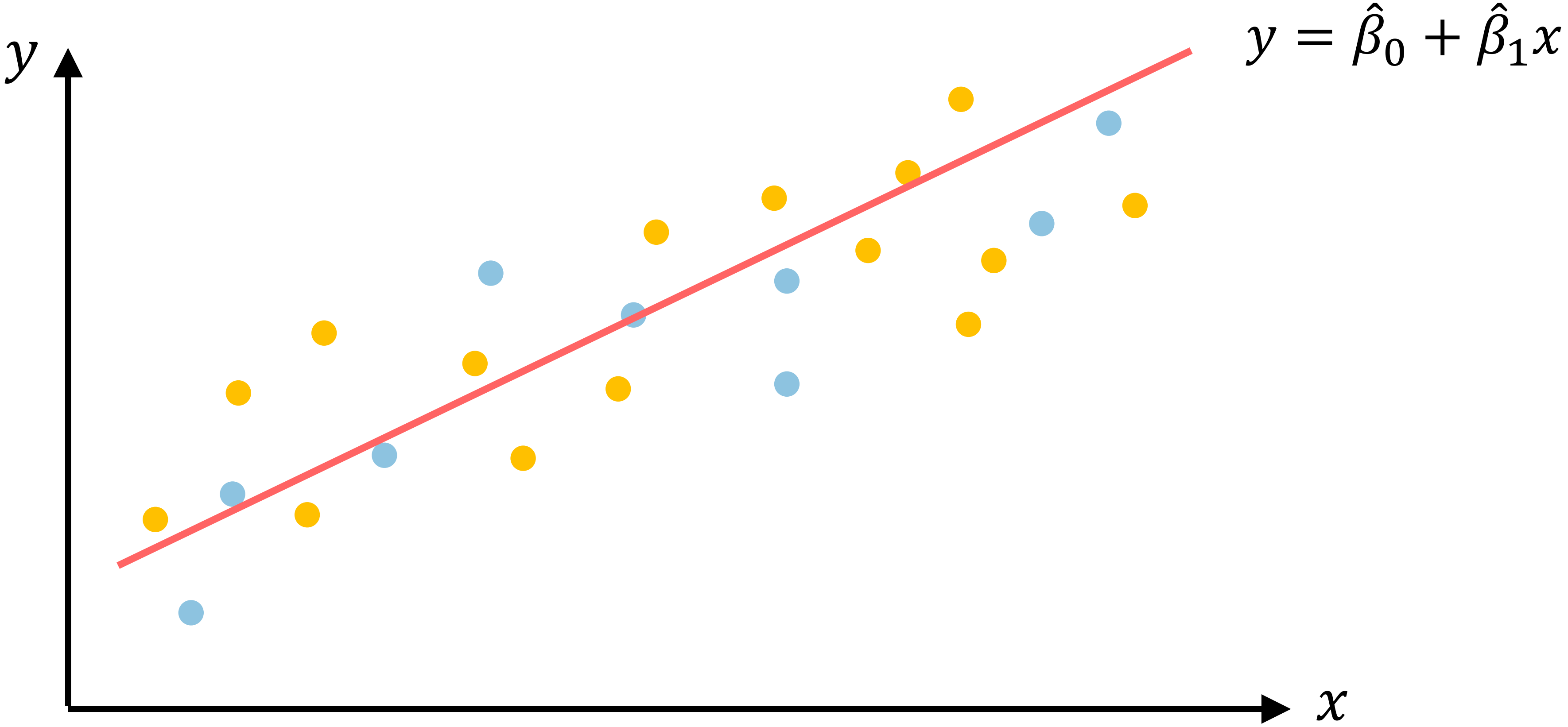
Linear Regression: Testing

- **Goal**: assess accuracy of learned linear model $y = \hat{\beta}_0 + \hat{\beta}_1 x$ using the test dataset



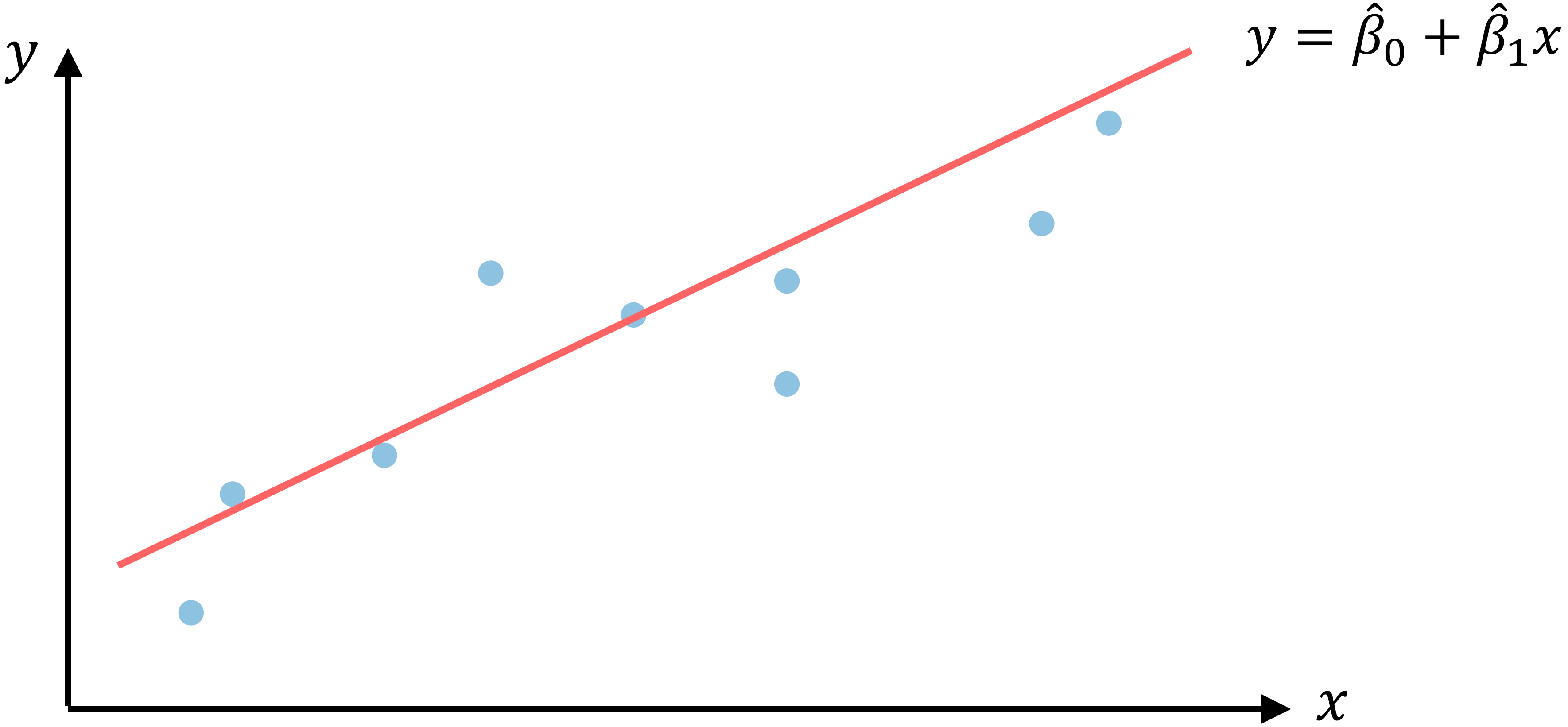
Linear Regression: Testing

- **Goal**: assess accuracy of learned linear model $y = \hat{\beta}_0 + \hat{\beta}_1 x$ using the test dataset



Linear Regression: Testing

- **Goal**: assess accuracy of learned linear model $y = \hat{\beta}_0 + \hat{\beta}_1 x$ using the test dataset



Linear Regression: Testing

- **Goal**: assess accuracy of learned linear model $y = \hat{\beta}_0 + \hat{\beta}_1 x$ using the test dataset

- Two metrics
  1. Mean Squared Error
  2. R^2 Coefficient of Determination

Mean Squared Error

Mean Squared Error (MSE) measures the average squared difference between the predicted values and the actual values.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y^i - \hat{y}^i\right)^2$$

n: number of predicted data points in the **test** data set

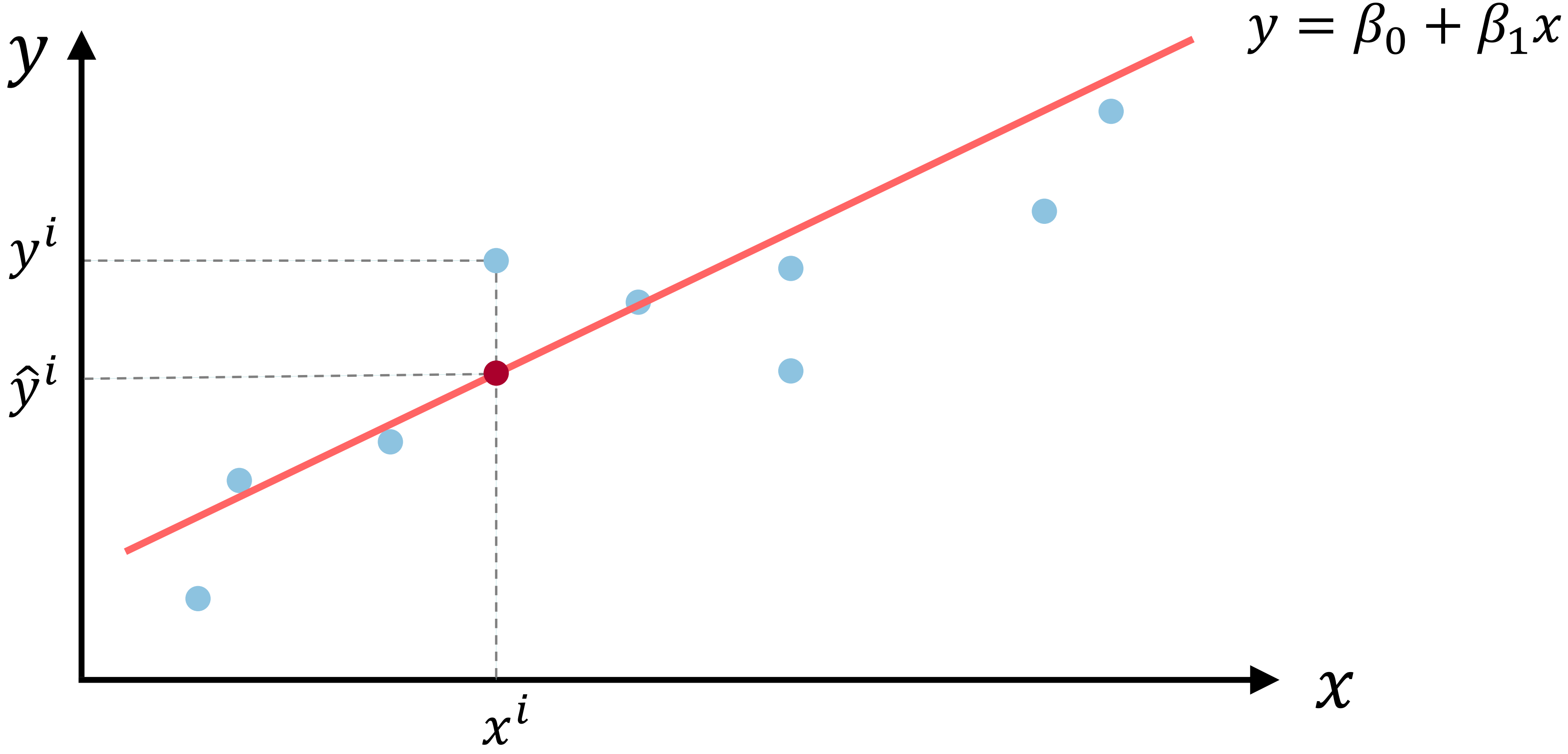
y^i: actual value in the **test** data set

$\hat{y}^i$: predicted value obtained using the hypothesis and x^i in the **test** data set.

Wikipedia

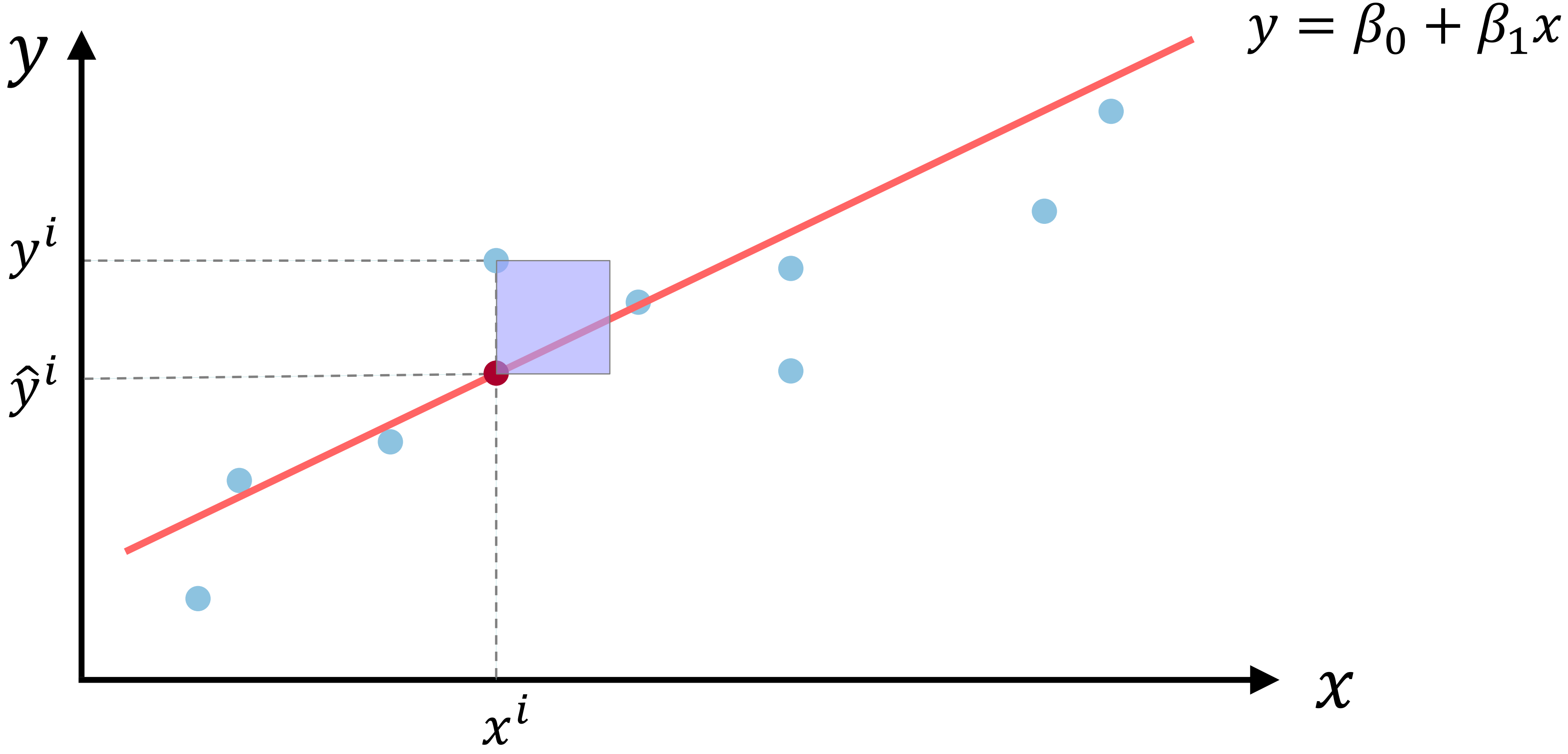
Mean Squared Error

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y^i - \hat{y}^i\right)^2$$



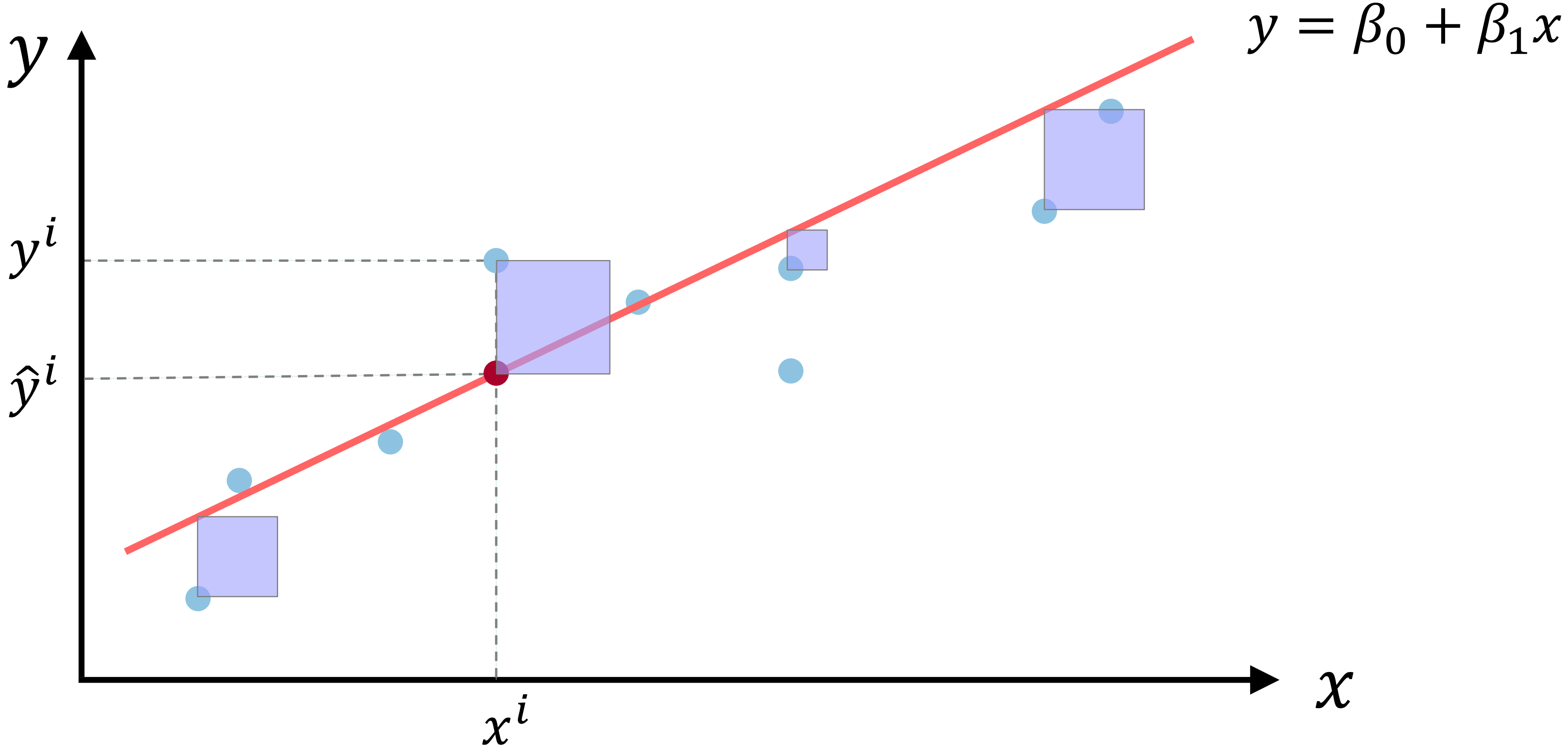
Mean Squared Error

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y^i - \hat{y}^i\right)^2$$



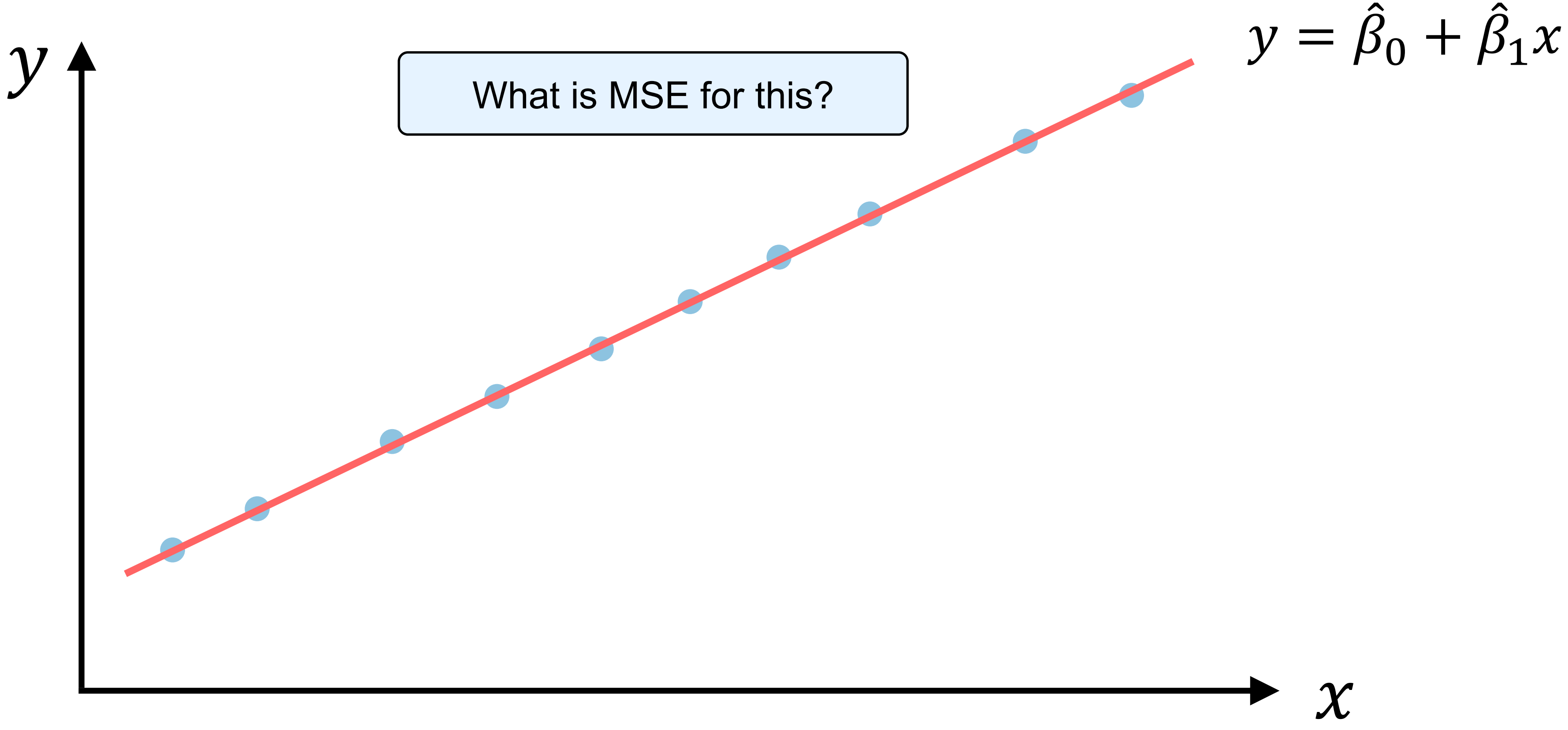
Mean Squared Error

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y^i - \hat{y}^i\right)^2$$



Mean Squared Error

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y^i - \hat{y}^i\right)^2$$



Mean Squared Error vs Cost Function

Mean Squared Error

What are the differences?

Cost Function

$$MSE = \frac{1}{\boxed{n}} \sum_{i=1}^{\boxed{n}} \left(y^i - \hat{y}^i\right)^2$$

$$J(\beta_0, \beta_1) = \frac{1}{2\boxed{m}} \sum_{i=1}^{\boxed{m}} \left(y^i - \hat{y}^i\right)^2$$

n is the number of predicted data points in the ***test*** dataset

used for **testing** a model

m is the number of predicted data points in the ***training*** dataset

used for **training** a model

R^2 Coefficient of Determination

R2 coefficient of determination is the proportion of the variation in the dependent variable that is predictable from the independent variable.

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

Residual sum of squares

Total sum of squares

where $$ss_{\text{res}} = \sum_{i=1}^{n} \left(y^i - \hat{y}^i\right)^2$$

$$ss_{\text{tot}} = \sum_{i=1}^{n} \left(y^i - \bar{y}\right)^2$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Wikipedia

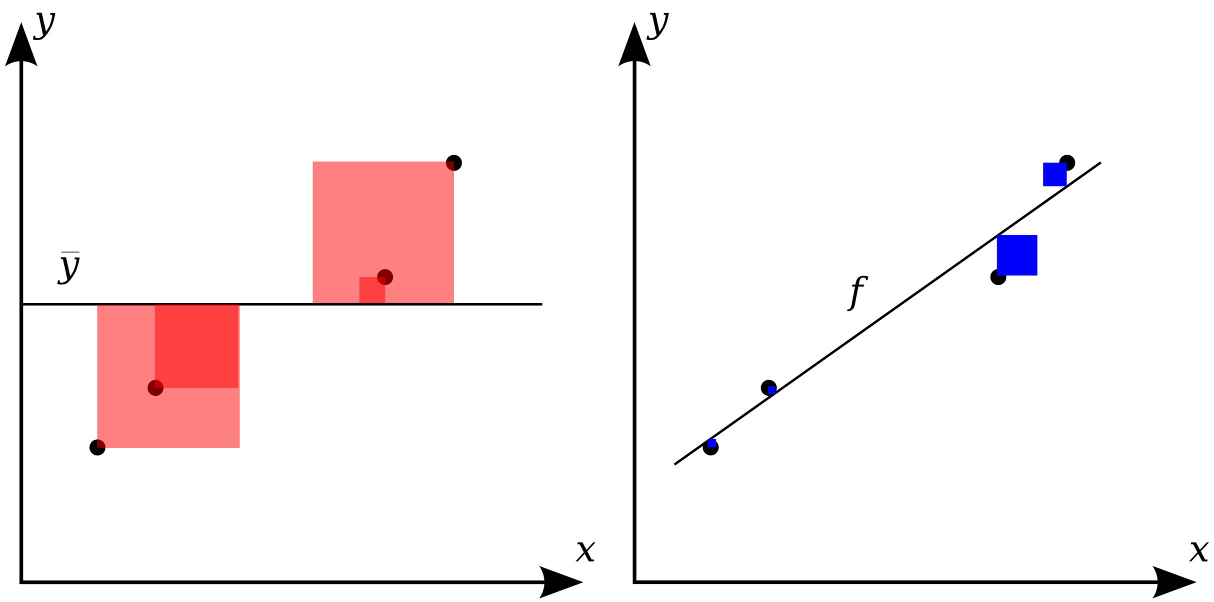
R^2 Coefficient of Determination

R2 coefficient of determination is the proportion of the variation in the dependent variable that is predictable from the independent variable.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

where $SS_{res} = \sum_{i=1}^{n} (y^i - \hat{y}^i)^2$

$$SS_{tot} = \sum_{i=1}^{n} (y^i - \bar{y})^2$$

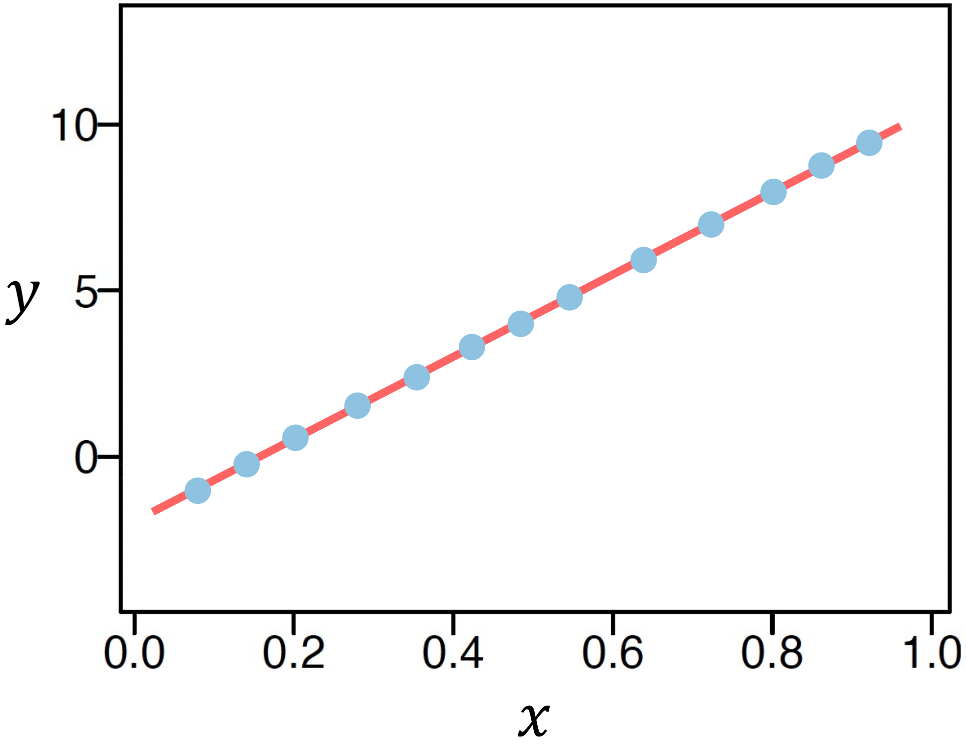


Wikipedia

R^2 Coefficient of Determination

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} = 1 - \frac{\sum_{i=1}^{n}\left(y^i - \hat{y}^i\right)^2}{\sum_{i=1}^{n}\left(y^i - \bar{y}\right)^2}$$

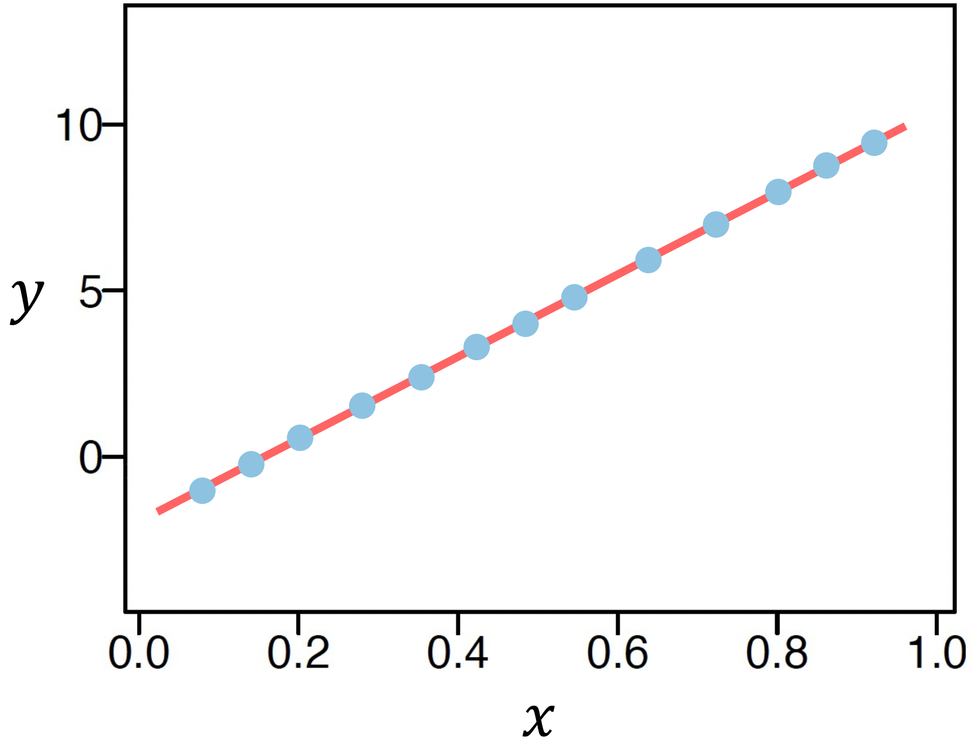
What is the R^2 value?



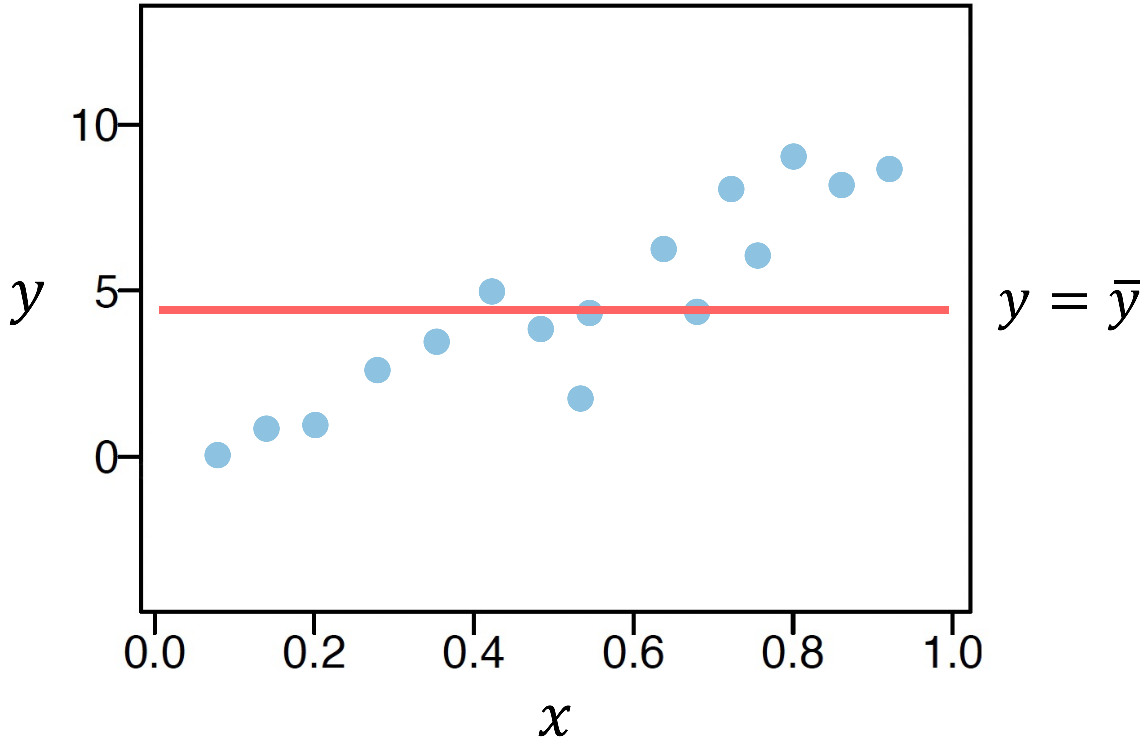
R^2 Coefficient of Determination

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} = 1 - \frac{\sum_{i=1}^{n}\left(y^i - \hat{y}^i\right)^2}{\sum_{i=1}^{n}\left(y^i - \bar{y}\right)^2}$$

$R^2 = 1.0$



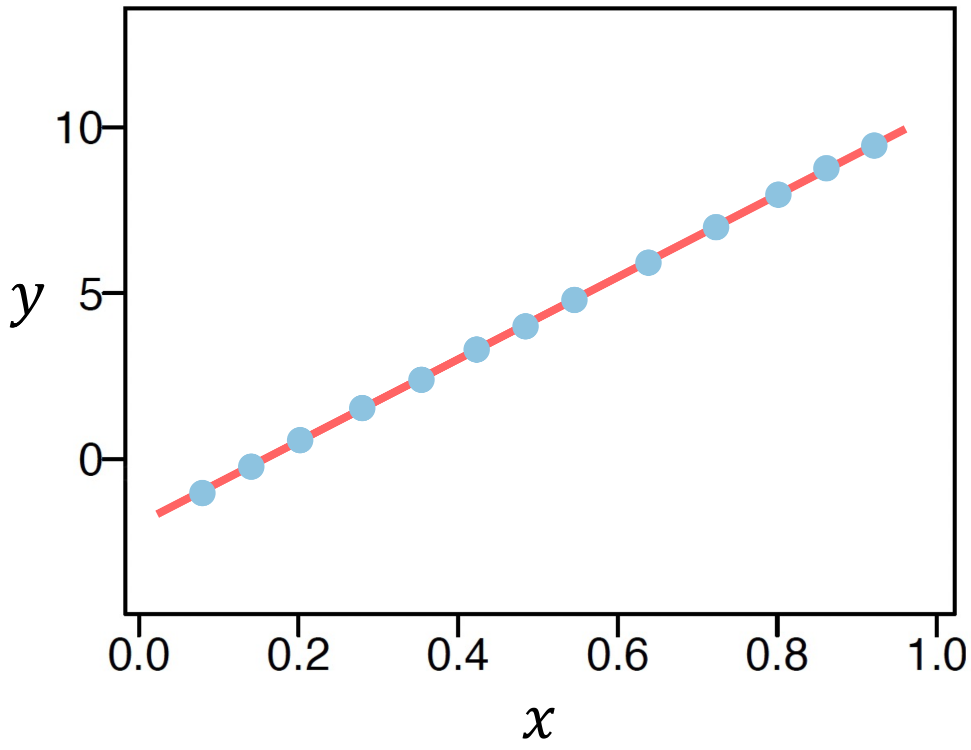
What is the R^2 value?



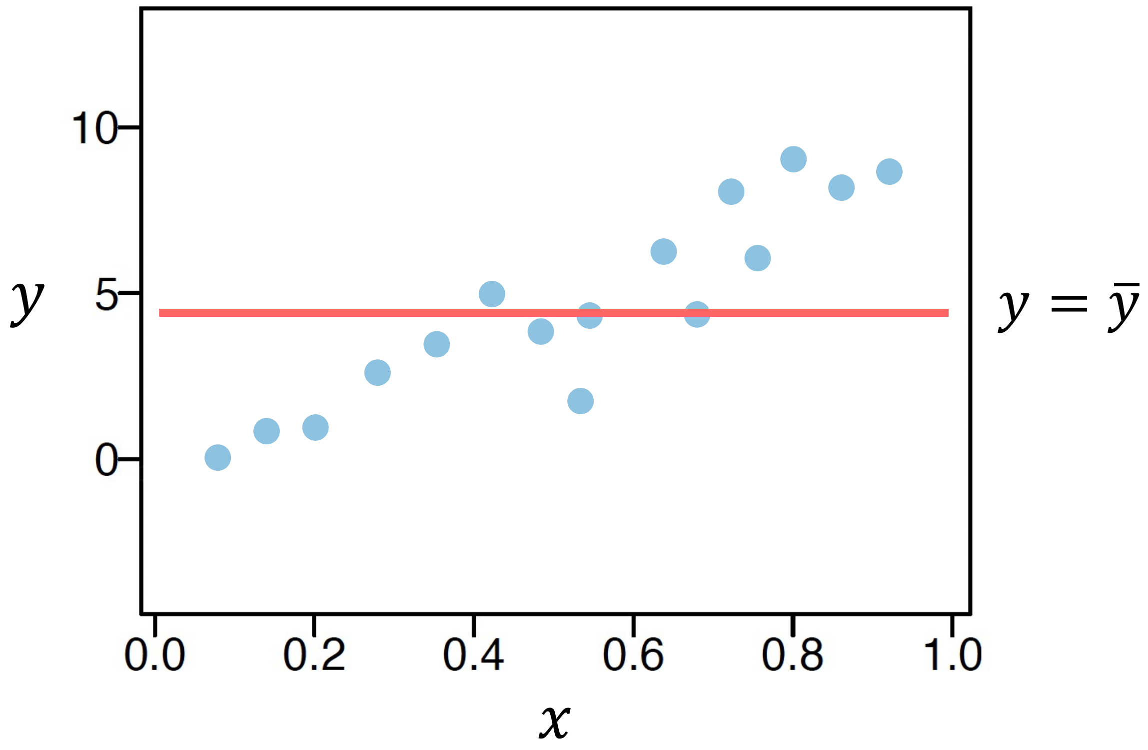
R^2 Coefficient of Determination

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} = 1 - \frac{\sum_{i=1}^{n}\left(y^i - \hat{y}^i\right)^2}{\sum_{i=1}^{n}(y^i - \bar{y})^2}$$

$R^2 = 1.0$

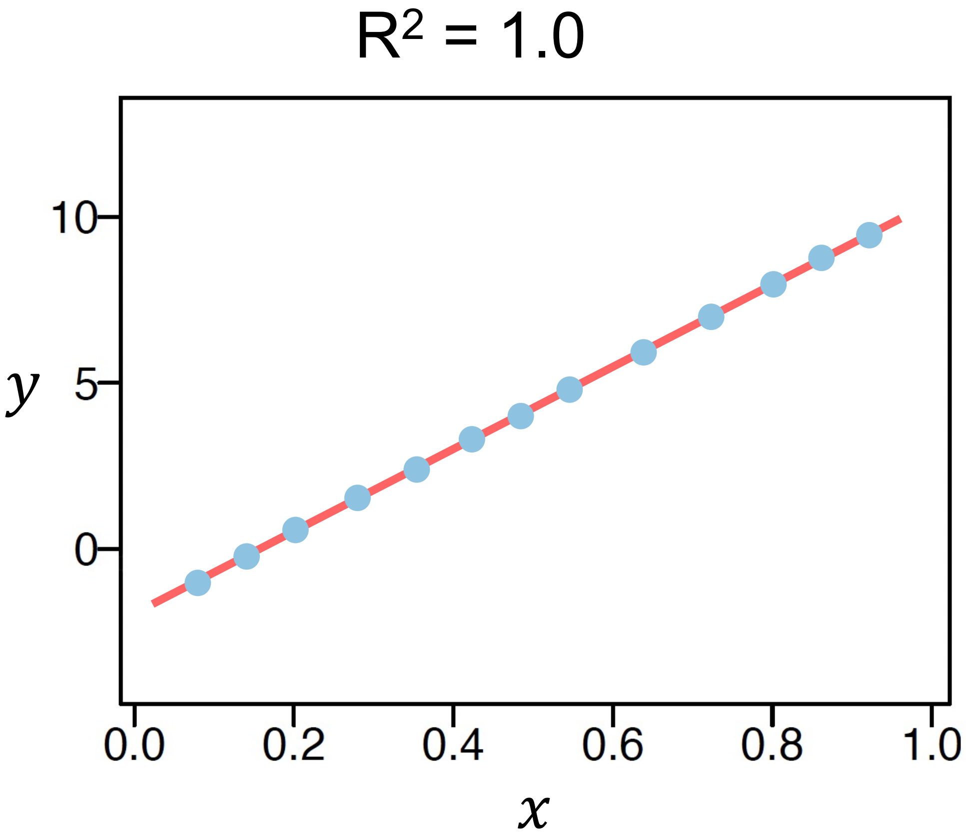


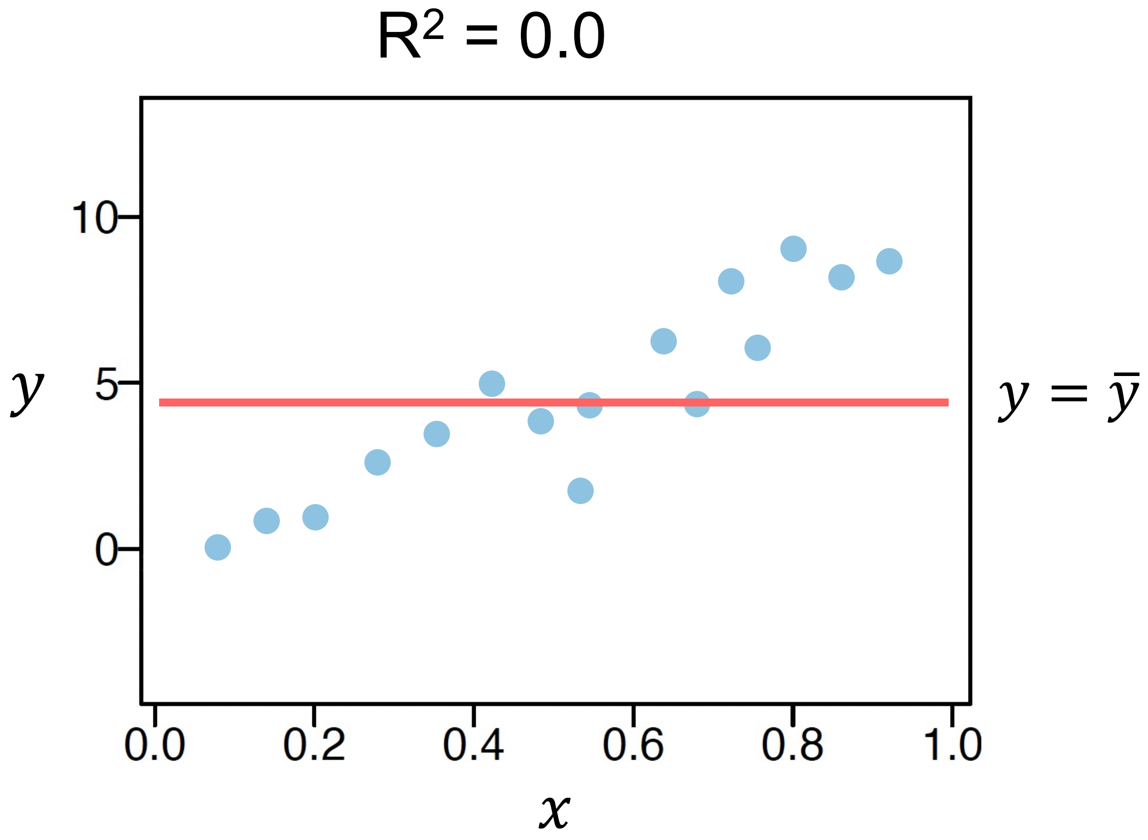
$R^2 = 0.0$



R^2 Coefficient of Determination

- R^2 = 1.0 when the predicted values match the observed values exactly
- R^2 = 0.0 when we choose the baseline model that always predict $\bar{y}$

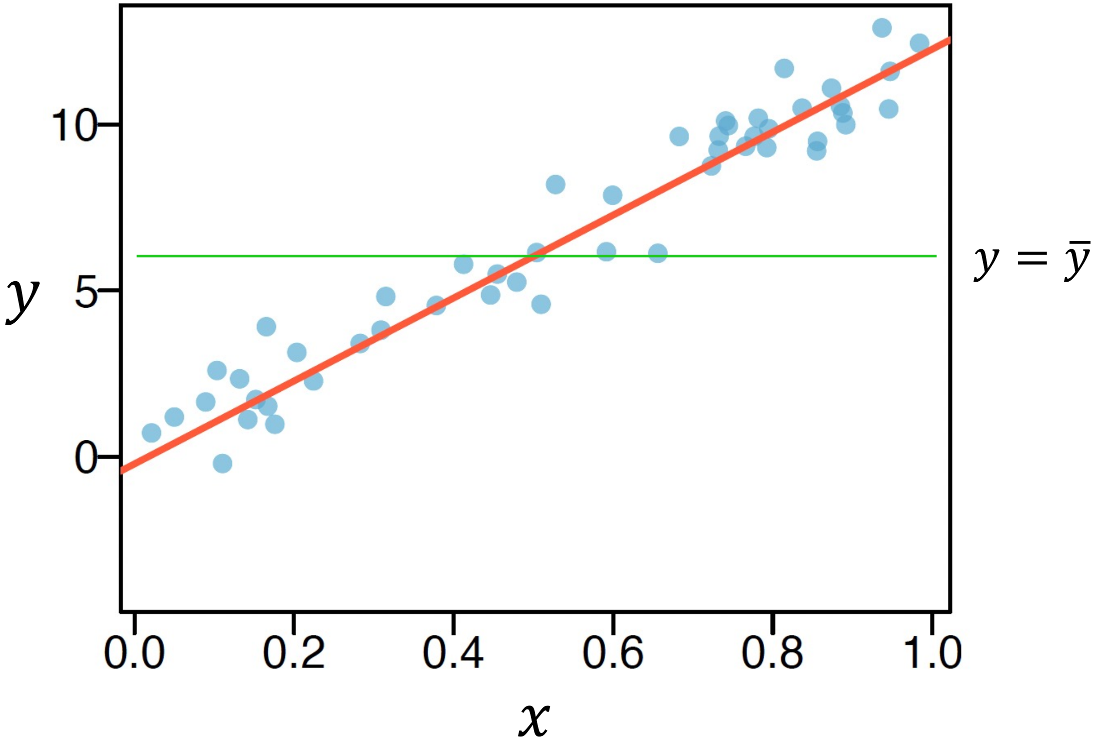


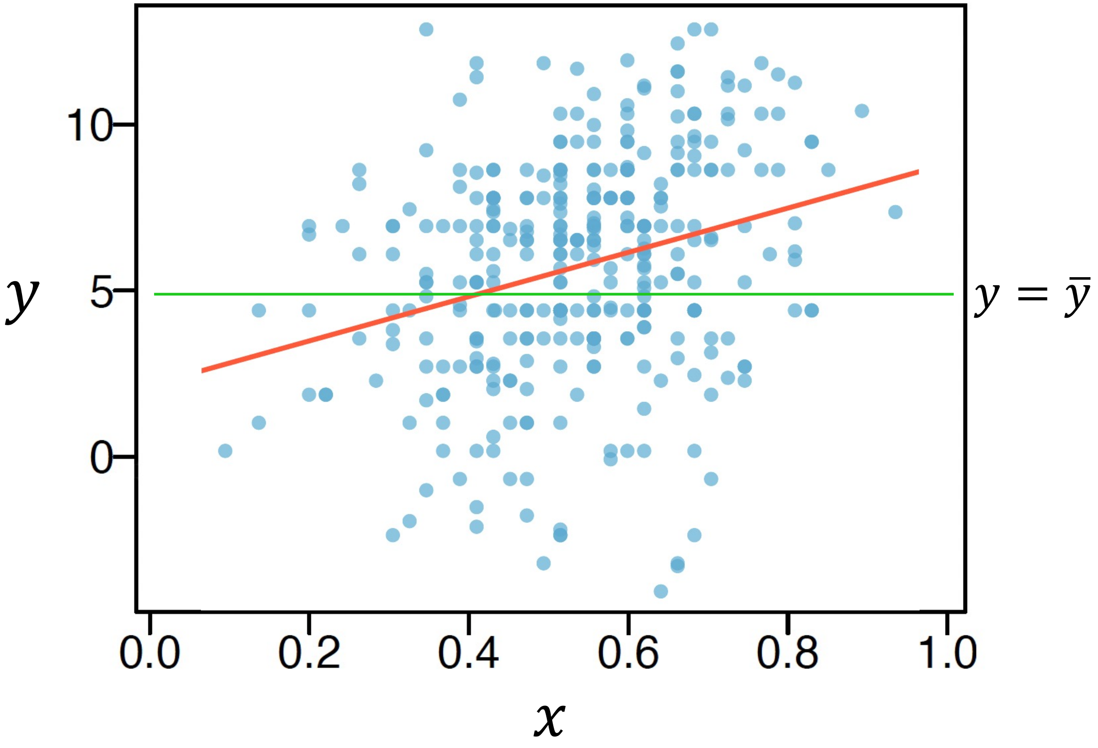


R^2 Coefficient of Determination

$$R^2 = 1 - \frac{ss_{\text{res}}}{ss_{\text{tot}}} = 1 - \frac{\sum_{i=1}^{n}\left(y^i - \hat{y}^i\right)^2}{\sum_{i=1}^{n}(y^i - \bar{y})^2}$$

Which figure has a larger R2 value?

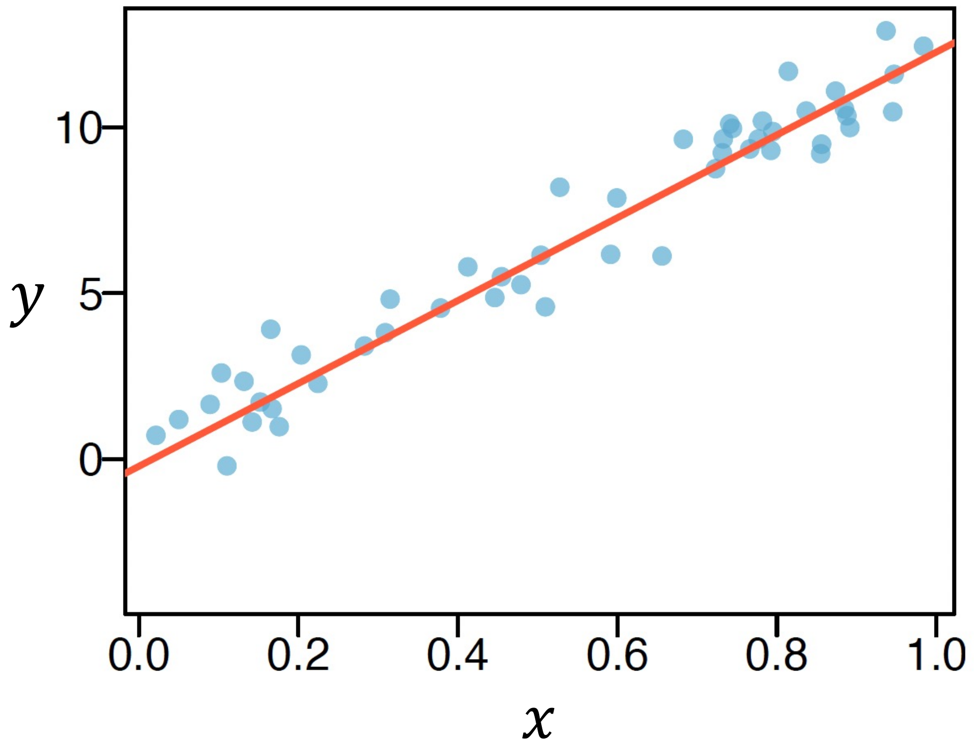




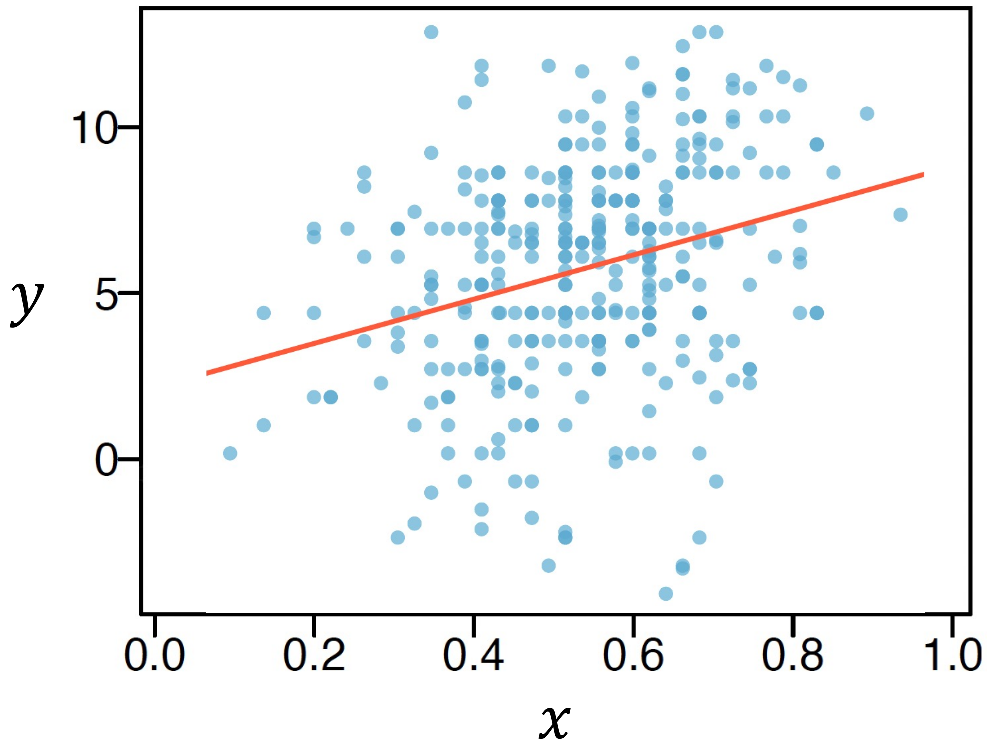
R^2 Coefficient of Determination

- R^2 is close to 1.0 if the test data is close to the learned linear function
- R^2 is close to 0.0 if there is no correlation between x and y

$R^2 = 0.91$



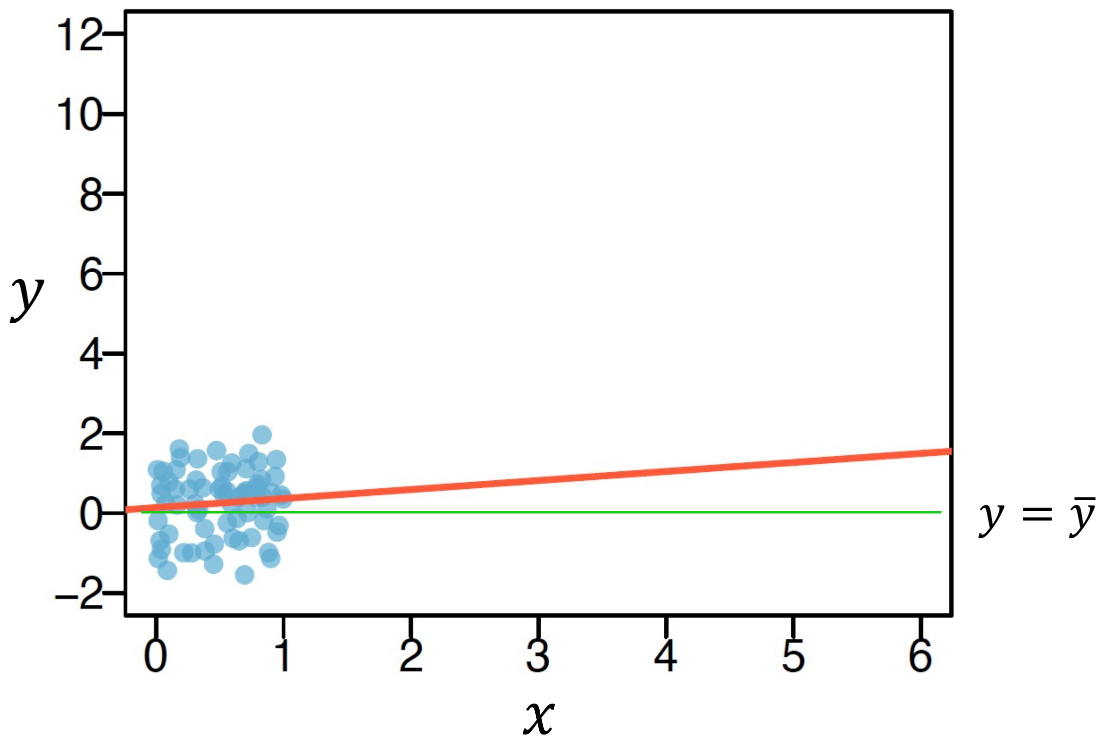
$R^2 = 0.28$

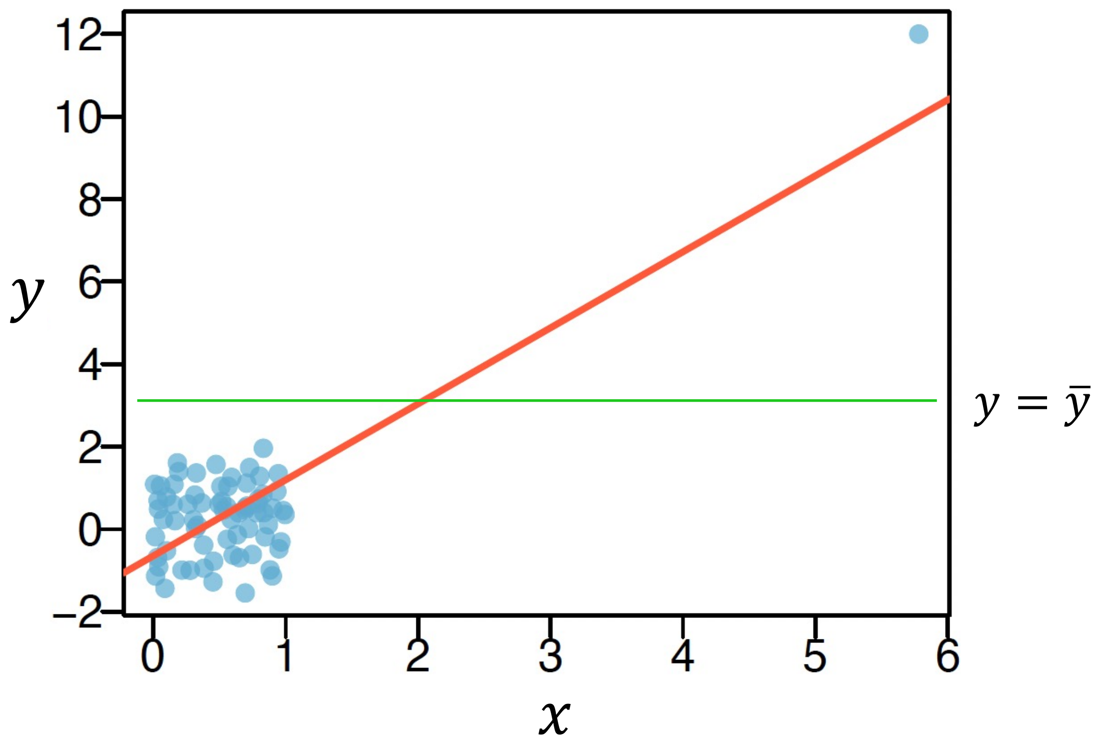


R^2 Coefficient of Determination

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} = 1 - \frac{\sum_{i=1}^{n} \left(y^i - \hat{y}^i\right)^2}{\sum_{i=1}^{n} (y^i - \bar{y})^2}$$

Which figure has a larger R2 value?

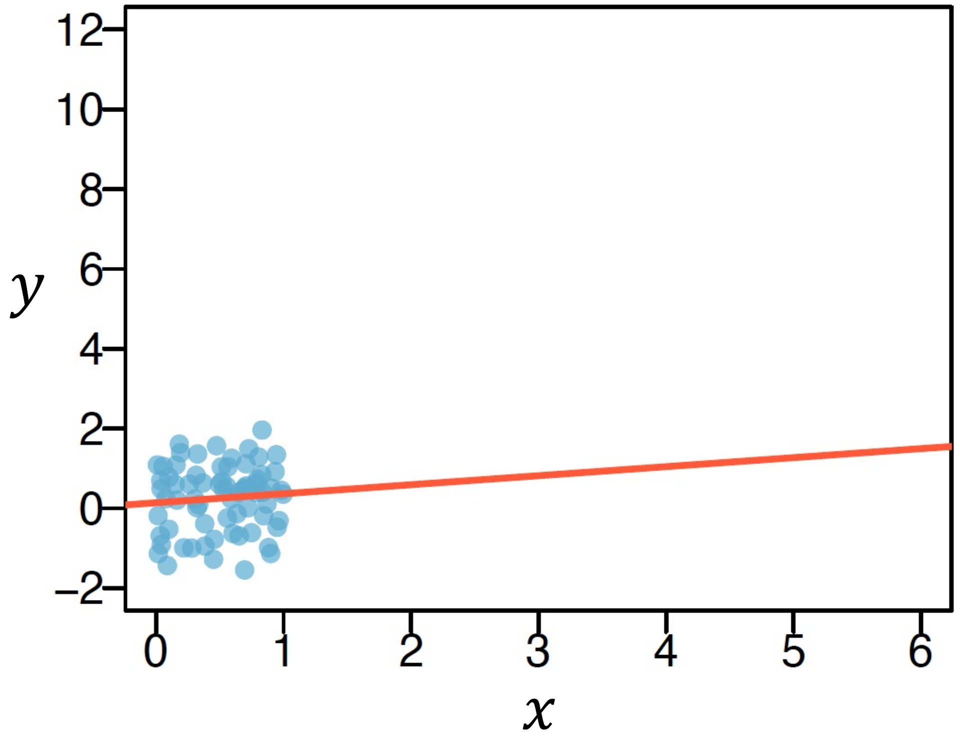




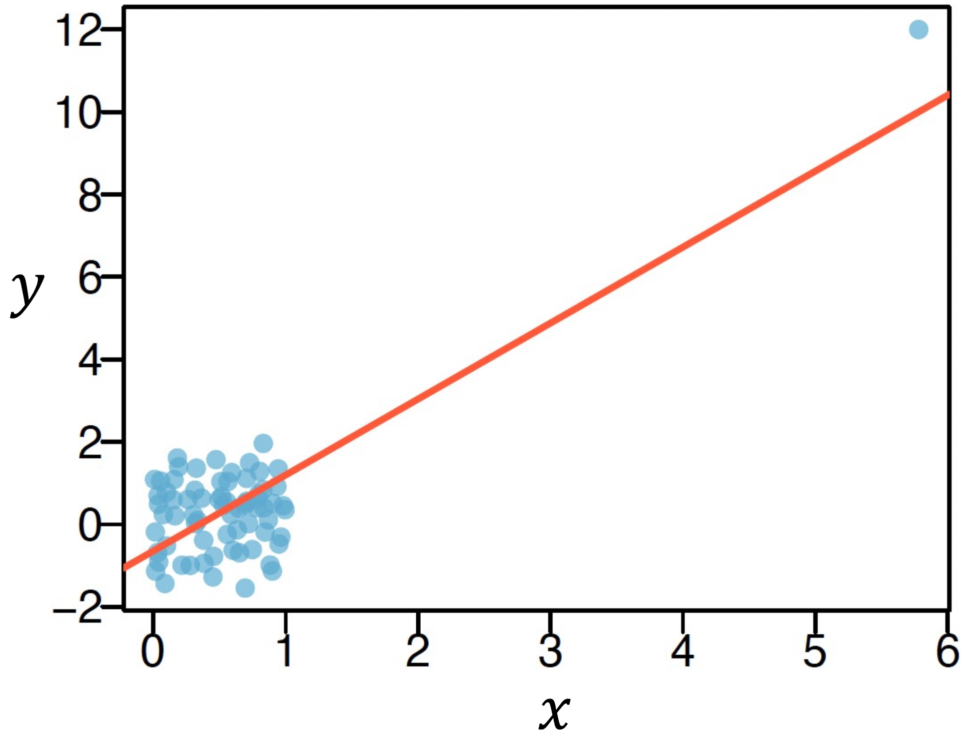
R^2 Coefficient of Determination

- Data **preprocessing** (e.g., removing outliers) and **visualization** are useful for training and testing a machine learning model.

$R^2 = 0.0064$



$R^2 = 0.627$



R2 vs MSE

What is the relation between R^2 and MSE?

$$R^2 = 1 - \frac{\boxed{\sum_{i=1}^{n}\left(y^i - \hat{y}^i\right)^2}}{\sum_{i=1}^{n}(y^i - \bar{y})^2}$$

$$MSE = \frac{1}{n}\boxed{\sum_{i=1}^{n}\left(y^i - \hat{y}^i\right)^2}$$

R2 vs MSE

What is the relation between R^2 and MSE?

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y^i - \hat{y}^i\right)^2}{\sum_{i=1}^{n}(y^i - \bar{y})^2} = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}\left(y^i - \hat{y}^i\right)^2}{\frac{1}{n}\sum_{i=1}^{n}(y^i - \bar{y})^2}$$

$$= 1 - \frac{\mathrm{MSE}}{\frac{1}{n}\sum_{i=1}^{n}(y^i - \bar{y})^2}$$

$$= 1 - \frac{\mathrm{MSE}}{C}$$ where $C = \frac{1}{n}\sum_{i=1}^{n}\left(y^i - \bar{y}\right)^2$

R2 vs MSE

Answer: R^2 is a rescaling of MSE to range $[0, 1]$

$$R^2 = 1 - \frac{\mathrm{MSE}}{C} \qquad \text{where } C = \frac{1}{n}\sum_{i=1}^{n}\left(y^i - \bar{y}\right)^2$$

Cohort Problem CS3

CS3. *Predict:* Write two functions

`def predict_norm(X, beta)` and

`def predict(df_feature, beta)`

that calculate the straight line equation given the features and its coefficient.

Hypothesis (learned model): $\hat{\boldsymbol{y}} = \boldsymbol{X} \times \boldsymbol{b}$

Cohort Problem CS4

CS4. *Splitting data:* Do the following tasks:

- Read RM as the feature and MEDV as the target from the data frame.
- Use Week 9's function `split_data()` to split the data into train and test using random_state=100 and test_size=0.3.
- Normalize and prepare the features and the target.
- Use the training data set and call `gradient_descent()` to obtain the theta.
- Use the test data set to get the predicted values.

Read Data:	`get_features_targets()`
Split Data:	`split_data()`
Training model:	`gradient_descent()`
Use model:	`predict()`

Cohort Problem CS5

CS5. *R2 Coefficient of Determination:* Write a function to calculate the coefficient of determination as given by the following equations.

$$R^2 = 1 - \frac{\mathrm{SS}_{\mathrm{res}}}{\mathrm{SS}_{\mathrm{tot}}} = 1 - \frac{\sum_{i=1}^{n}\left(y^i - \hat{y}^i\right)^2}{\sum_{i=1}^{n}(y^i - \bar{y})^2}$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

Cohort Problem CS6

CS6. *Mean Squared Error:* Create a function to calculate the MSE as given below.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y^i - \hat{y}^i\right)^2$$

Thank You!