2. Supervised Learning: Simple Linear Regression

Linear Regression

Example

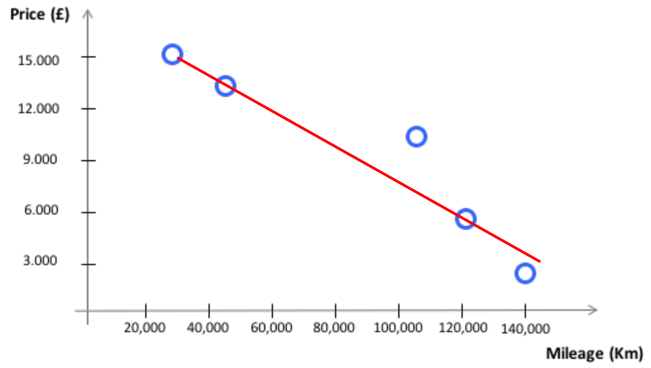
- Prediction of second-hand car prices based on features such as mileage
- It is a supervised learning problem because one has access to input-output examples
- It is a regression problem because the output variable is continuous-valued
- Input variables also known as features or regressors
- Output variable also known as response or prediction

Training set

Input Variable (Mileage, Km)	*Output Variable (Cost, £)*
25,000	*16,000*
105,000	*11,500*
120,000	*6,000*
140,000	*3,000*
45,000	*13,500*

Training sample

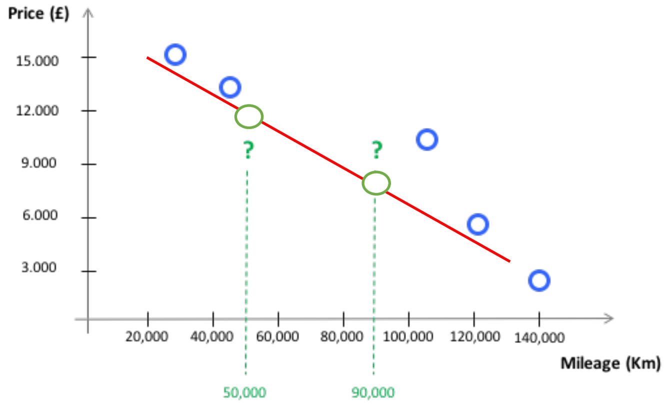
(x_i, y_i)



Testing set

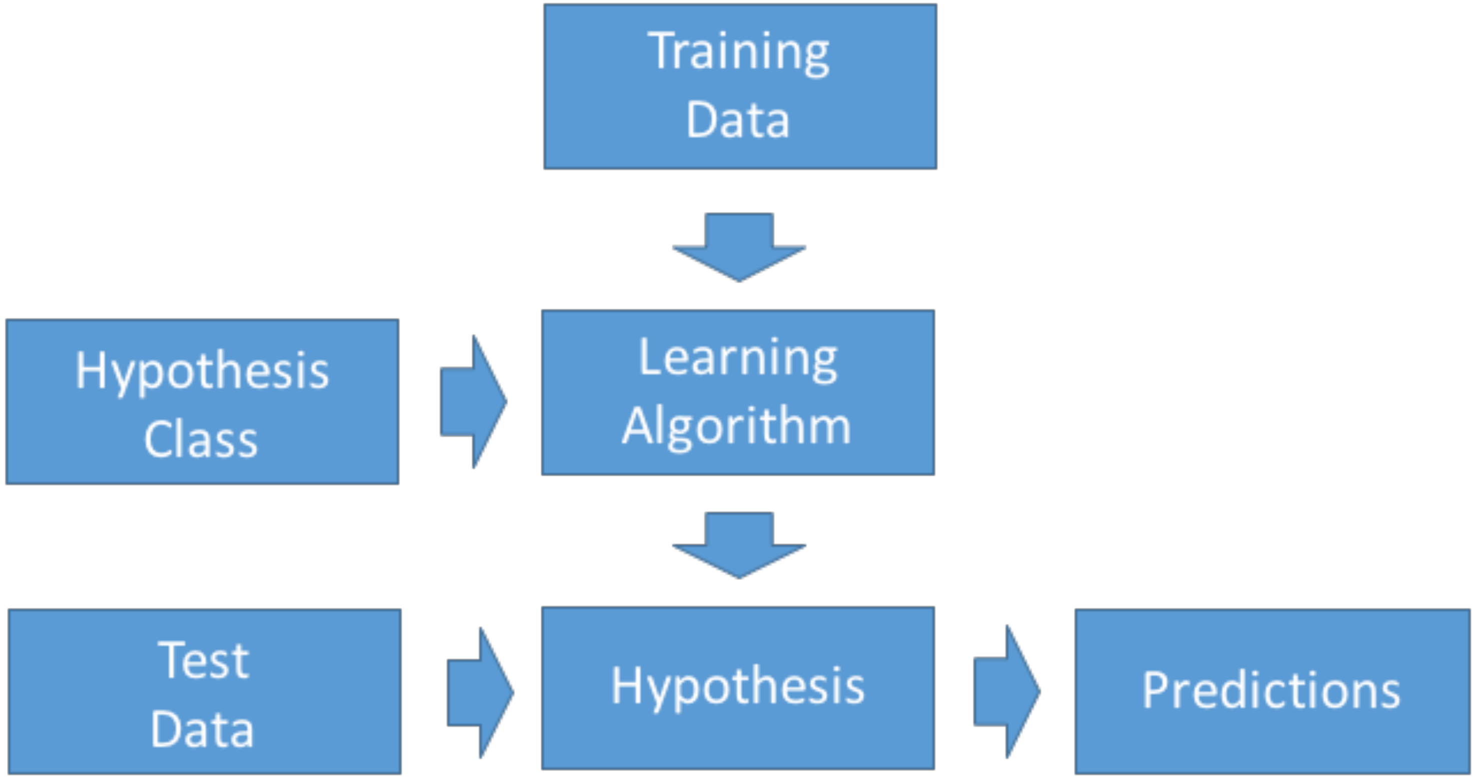
Input Variable (Mileage, Km)	*Output Variable (Cost, £)*
50,000	*?*
90,000	*?*

Test sample



Linear Regression: Approach

Process



- One is given access to a **training data** – consisting of various feature-response pairs – and **testing data** – consisting of features points with unknown response.
- One is also given a **hypothesis** (or **model**) **class** containing a series of **hypotheses** (or **models**) that potentially explain the relationship between the features and responses.
- The **learning algorithm** selects a **hypothesis** (or **model**) from the **hypothesis** (or **model**) **class** that fits the **training data**.
- Such selected **hypothesis** can then be used on the **testing data** to determine the response associated with the new features.

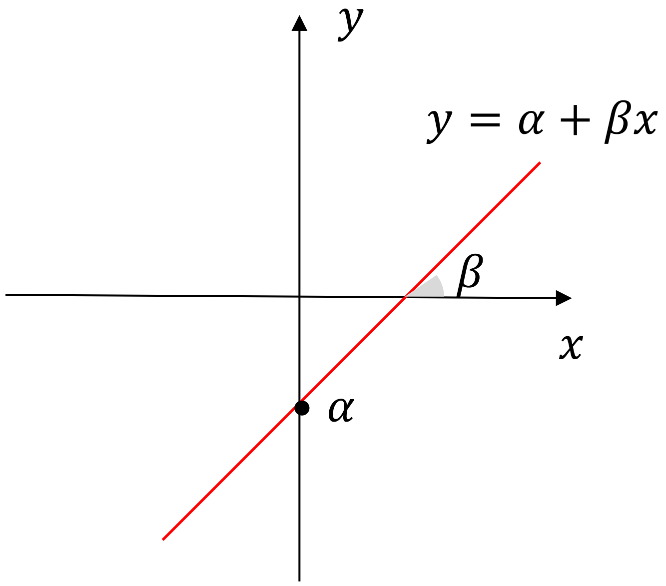
Linear Regression: Model

Linear Model

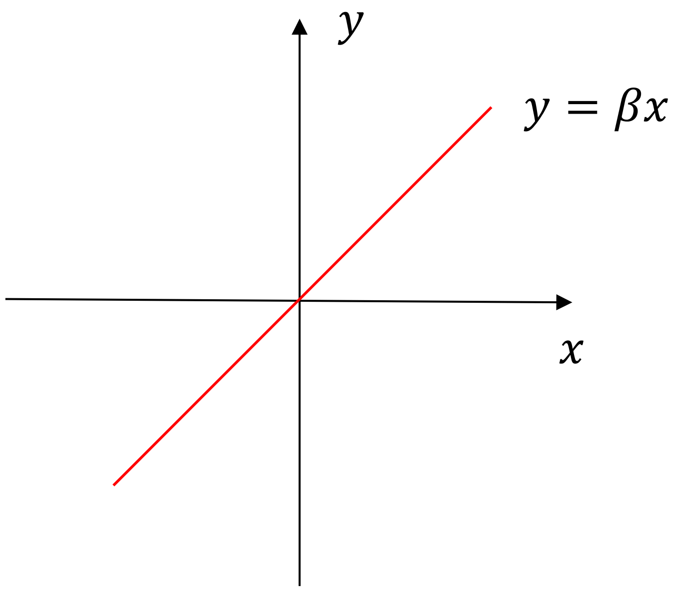
In linear regression, the relationship between the input and output variables is expressed as follows:

$$y = \alpha + \beta x$$

where α and β are model parameters



Examples



$\alpha = 0\,; \beta > 0$

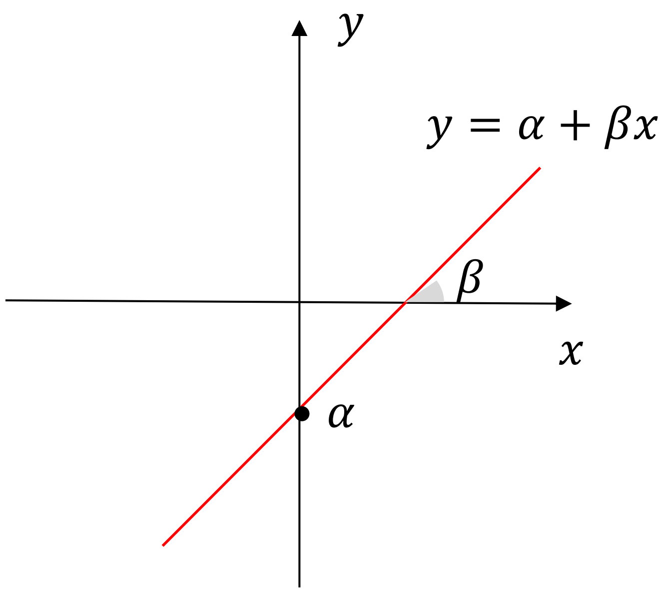
Linear Regression: Model

Linear Model

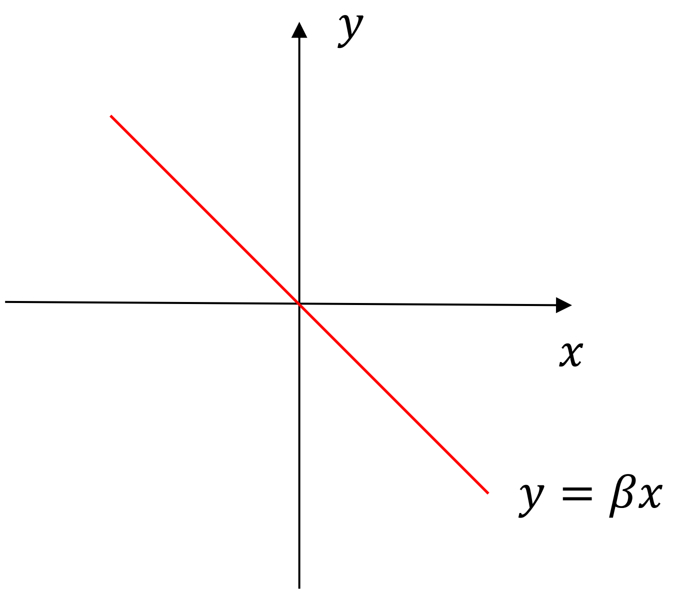
In linear regression, the relationship between the input and output variables is expressed as follows:

$$y = \alpha + \beta x$$

where α and β are model parameters



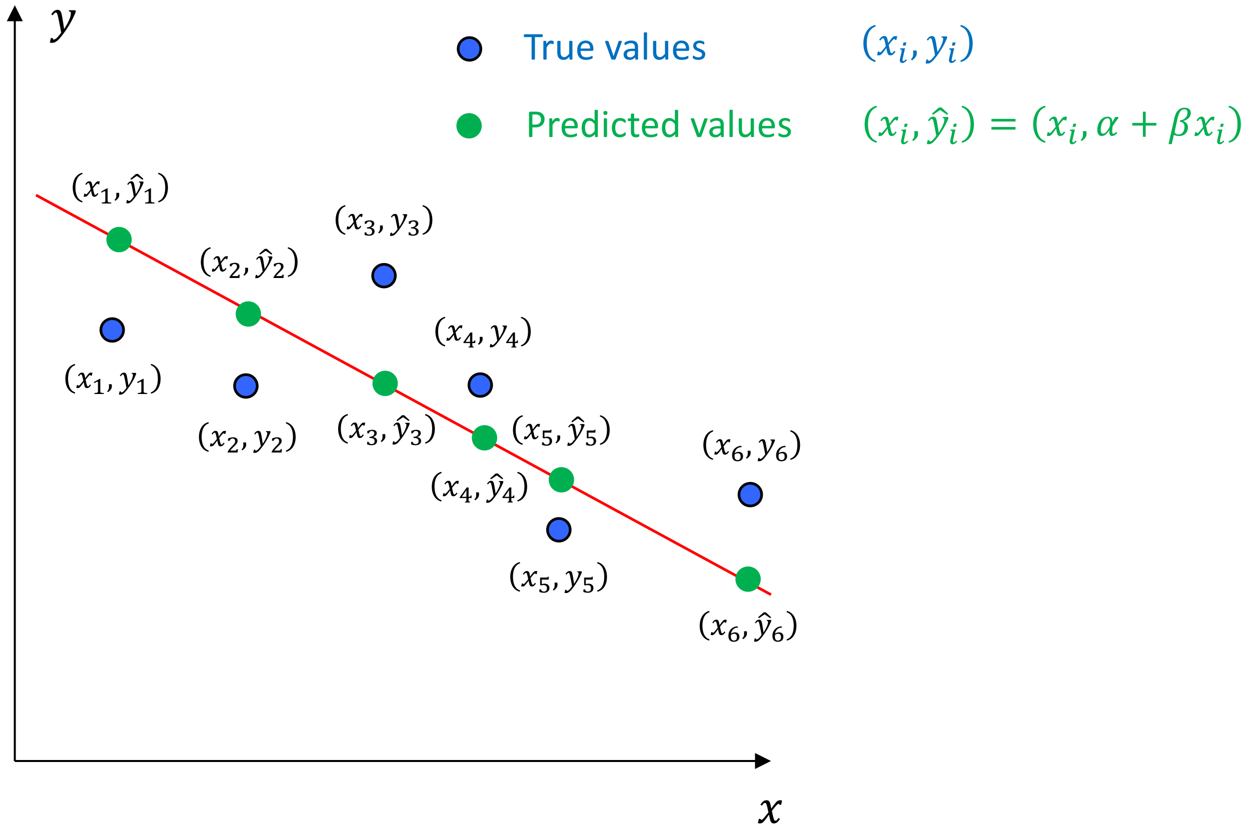
Examples



$$\alpha = 0\,; \beta < 0$$

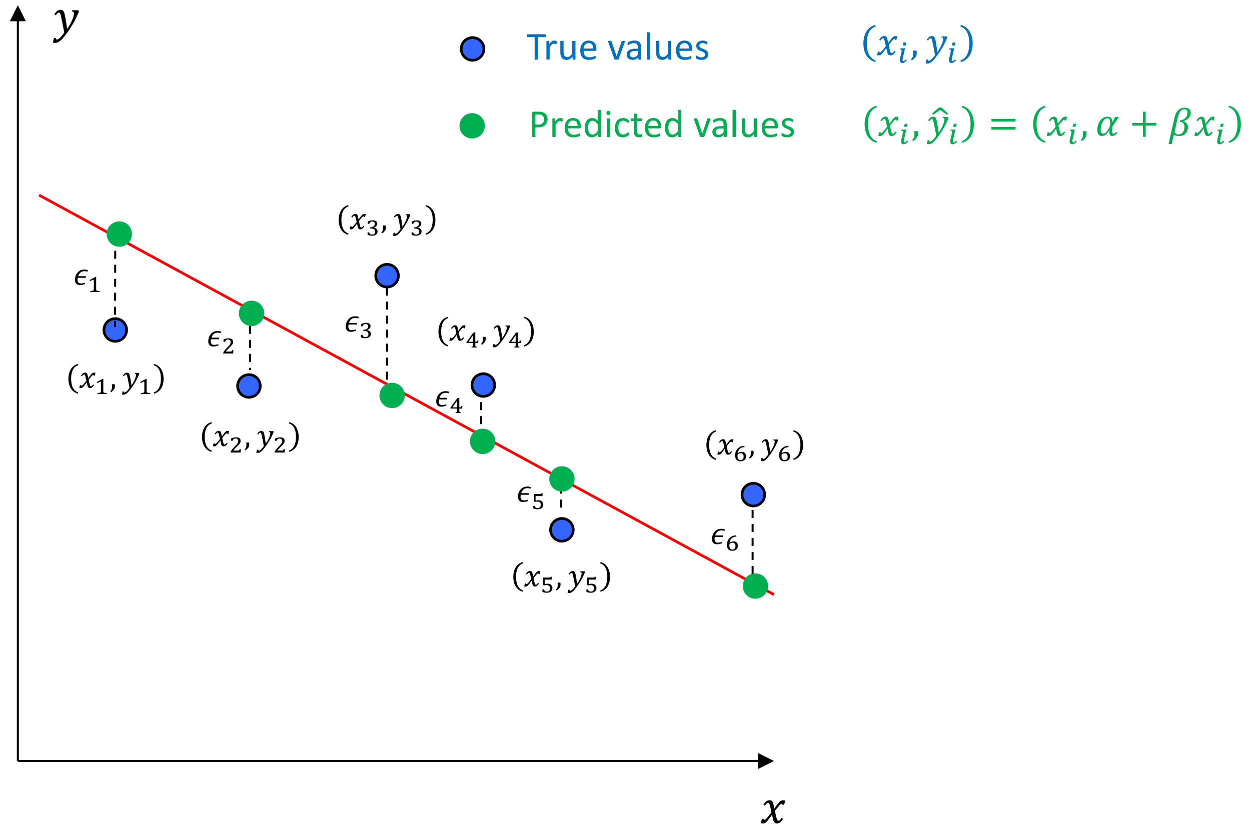
Linear Regression: Cost Function

How does the learning algorithm select the linear hypothesis / model? We need a cost function…



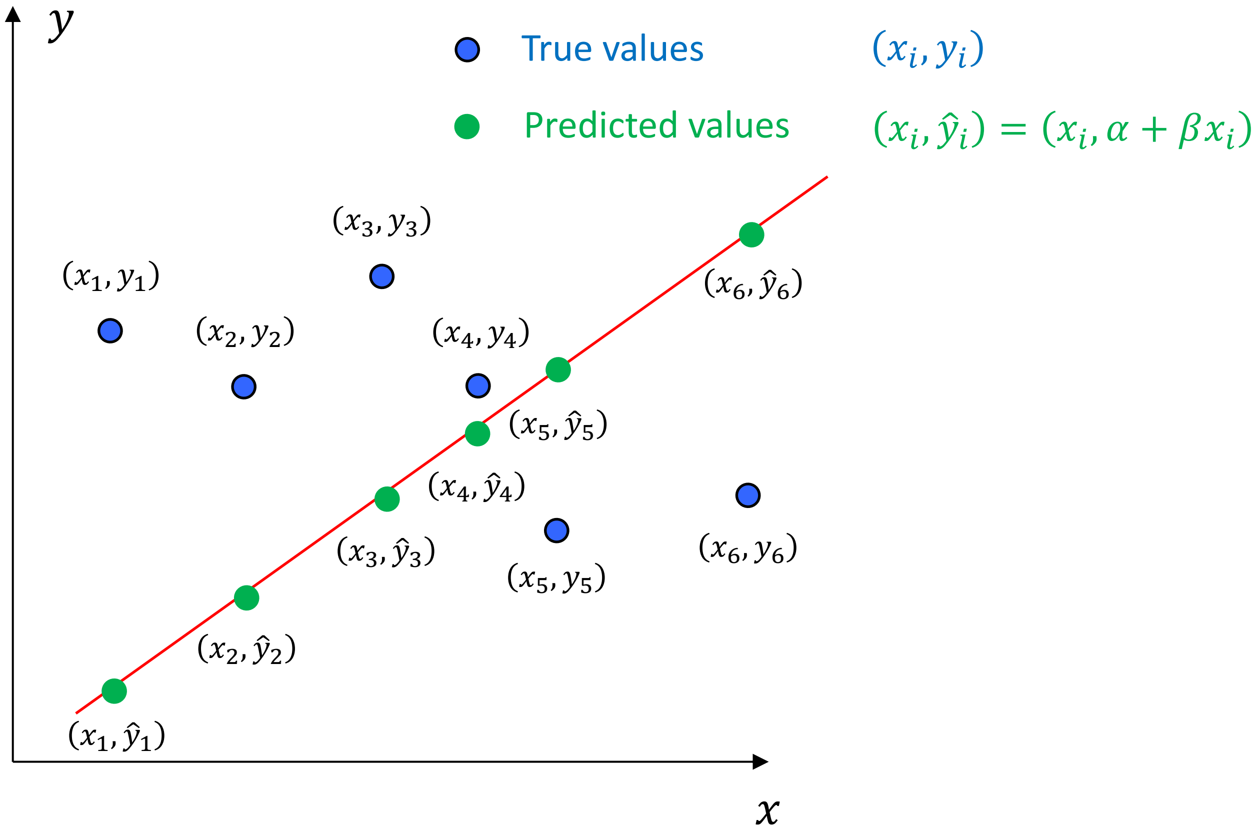
Linear Regression: Cost Function

How does the learning algorithm select the linear hypothesis / model? We need a cost function…



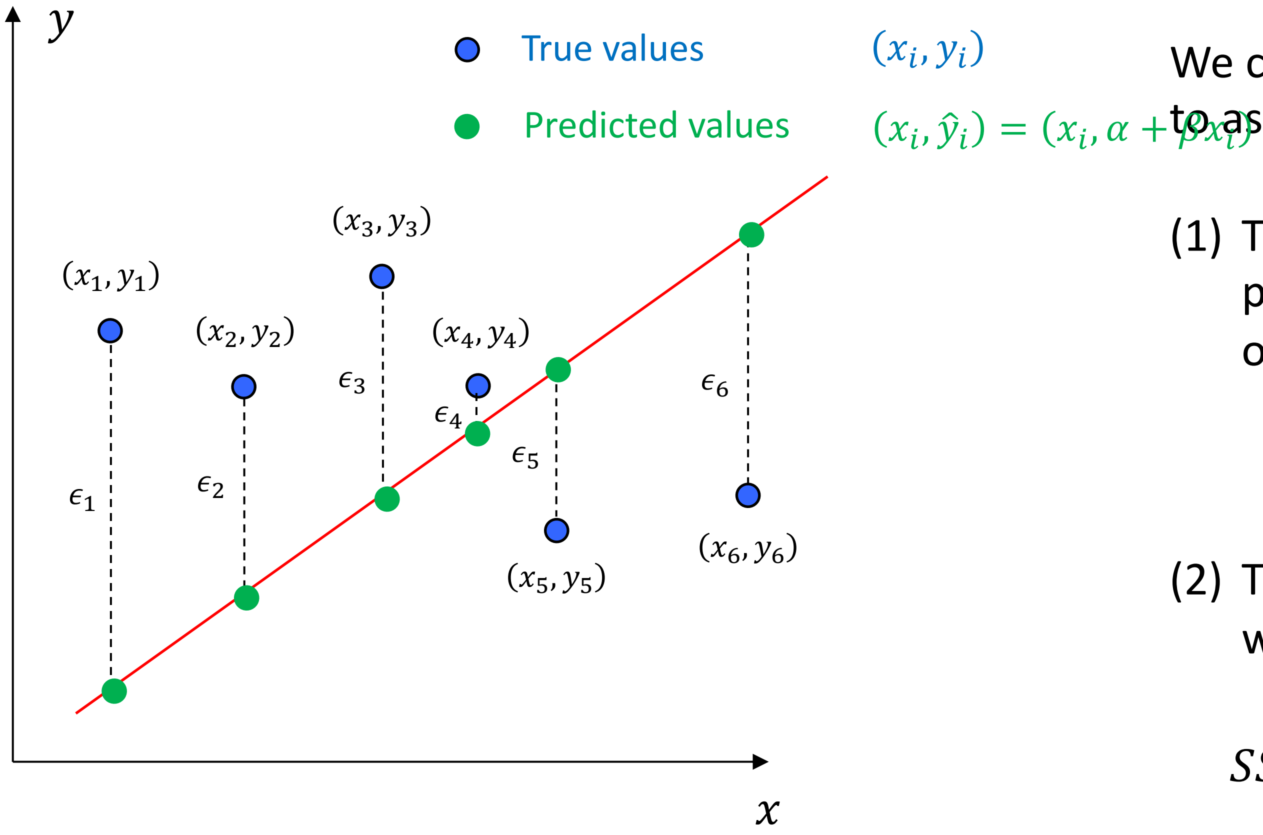
Linear Regression: Cost Function

How does the learning algorithm select the linear hypothesis / model? We need a cost function…



Linear Regression: Cost Function

How does the learning algorithm select the linear hypothesis / model? We need a cost function…



Cost Function – Sum-of-Squared Residuals

We can now define two quantities that can be used to assess the model fit to the data:

(1) The individual residual measures how much the predicted output value deviates from the true output value given the input value:

$$\epsilon_i = y_i - \hat{y}_i = y_i - (\alpha + \beta x_i)$$

(2) The sum of squared residuals measures how well overall the model fits the data:

$$SSR(\alpha, \beta) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} \big(y_i - (\alpha + \beta x_i)\big)^2$$

Linear Regression: Learning Algorithm

Learning Algorithm

We should select the linear regression model parameters that minimize the sum-of-squared residuals as follows:

$$(\alpha^*, \beta^*) = \underset{\alpha,\beta}{\operatorname{argmin}} \, SSR(\alpha, \beta) = \underset{\alpha,\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - (\alpha + \beta x_i)\right)^2$$

This leads to the optimal model parameters given by:

$$\alpha^* = \bar{y} - \beta^* \cdot \bar{x} \qquad \beta^* = \frac{\sum_{i=1}^{n} x_i (y_i - \bar{y})}{\sum_{i=1}^{n} x_i (x_i - \bar{x})}$$

where $\bar{x} = 1/n \cdot \sum_{i=1}^{n} x_i$ and $\bar{y} = 1/n \cdot \sum_{i=1}^{n} y_i$

New Predictions

The new response $\hat{y}$ associated with a new data point x is now given by:

$$\hat{y} = \alpha^* + \beta^* x$$