3. Supervised Learning: Multivariate Linear Regression

Multivariate Linear Regression

Example

- Prediction of second-hand car prices based on features such as mileage, # doors, # cylinders
- 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 Variables (Mileage, Km)	# doors	# cylinders	Output Variable (Cost, £)
25,000	3	6	16,000
105,000	5	12	11,500
120,000	5	8	6,000
140,000	3	8	3,000
45,000	3	12	13,500

Training sample

$(x_{i,1}, x_{i,2}, x_{i,3}, y_i)$

Testing set

Input Variables: Input Variable (Mileage, Km)	# doors	# cylinders	Output Variable (Cost, £)
50,000	50,000	50,000	?
90,000	90,000	90,000	?

Test sample

Multivariate Linear Regression

UCL

Example

- Prediction of second-hand car prices based on features such as mileage, # doors, # cylinders
- 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 Variables			*Output Variable*
(Mileage, Km)	*# doors*	*# cylinders*	*(Cost, £)*
25,000	*3*	*6*	*16,000*
105,000	*5*	*12*	*11,500*
120,000	*5*	*8*	*6,000*
140,000	*3*	*8*	*3,000*
45,000	*3*	*12*	*13,500*

Training sample

$(x_{i,1}, x_{i,2}, x_{i,3}, y_i)$

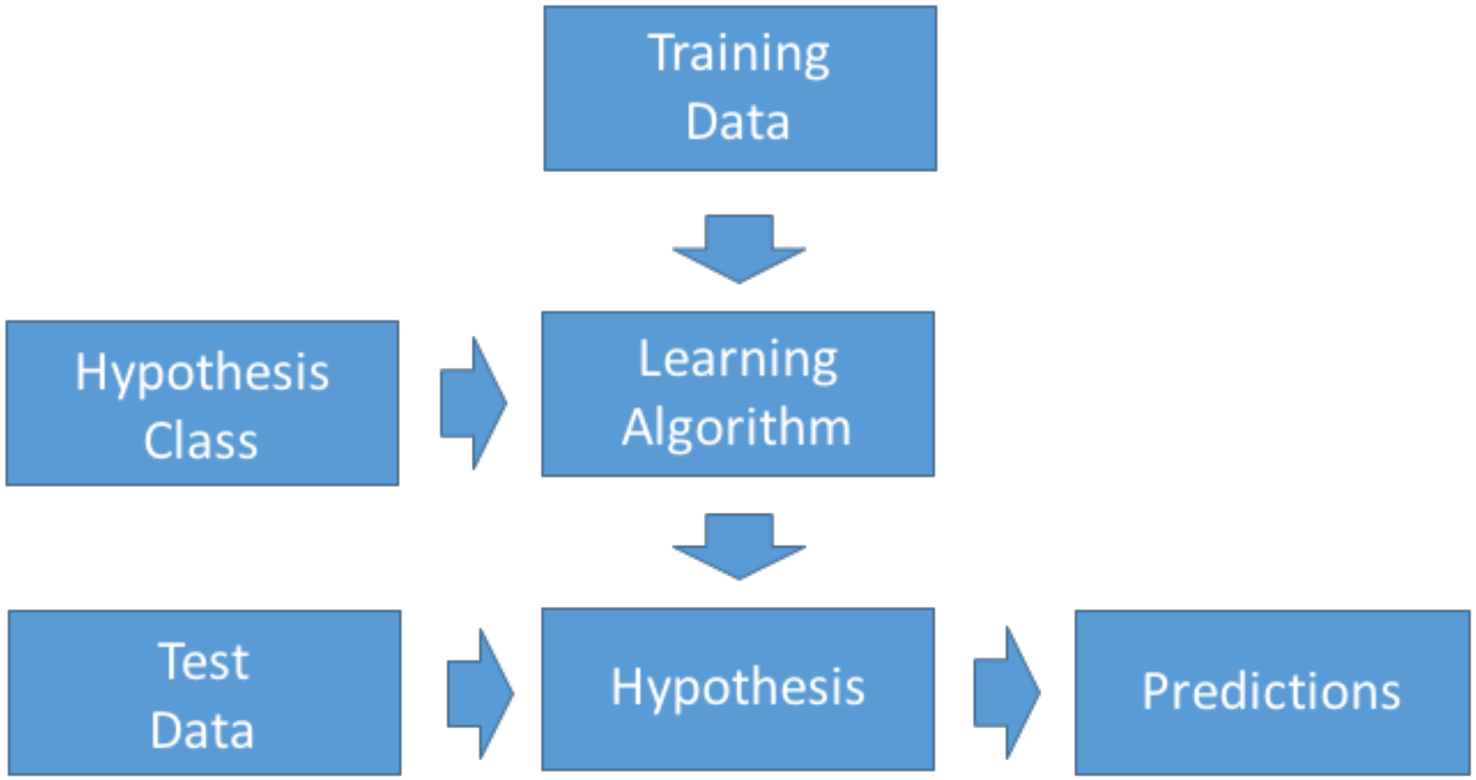
Testing set

Input Variables			*Output Variable*
Input Variable (Mileage, Km)	*# doors*	*# cylinders*	*(Cost, £)*
50,000	*50,000*	*50,000*	*?*
90,000	*90,000*	*90,000*	*?*

Test sample

Multivariate Linear Regression

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.

Multivariate Linear Regression: Model

Multivariate Linear Model

The relationship between the input and output variables can be expressed as follows:

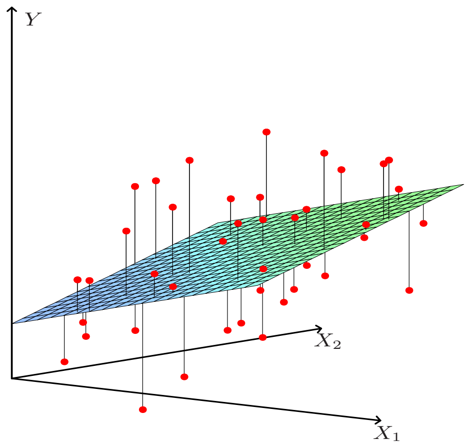
$$y_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_m x_{i,m} \quad (i = 1, \ldots, n)$$

where $\beta_j, j = 0, \ldots, m$ are the model parameters

The relationship between the input and output variables can also be expressed as follows:

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,m} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_m \end{bmatrix} = \boldsymbol{X\beta}$$

Geometrical Interpretation



Multivariate Linear Regression: Model

Multivariate Linear Model

The relationship between the input and output variables can be expressed as follows:

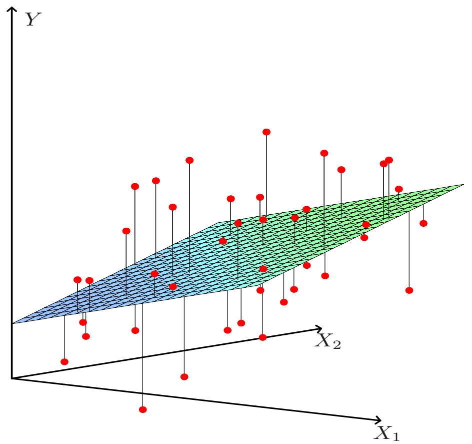
$$y_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_m x_{i,m} \quad (i = 1, \ldots, n)$$

where $\beta_j, j = 0, \ldots, m$ are the model parameters

The relationship between the input and output variables can also be expressed as follows:

$$\boldsymbol{y} = \begin{bmatrix} 16{,}000 \\ \vdots \\ 13{,}500 \end{bmatrix} = \begin{bmatrix} 1 & 25000 & 3 & 6 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 45{,}000 & 3 & 12 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_m \end{bmatrix} = \boldsymbol{X\beta}$$

Geometrical Interpretation



Input Variable (Mileage, Km)	# doors	# cylinders	Output Variable (Cost, £)
25,000	3	6	16,000
105,000	5	12	11,500
120,000	5	8	6,000
140,000	3	8	3,000
45,000	3	12	13,500

Multivariate Linear Regression: Model

Multivariate Linear Model

The relationship between the input and output variables can be expressed as follows:

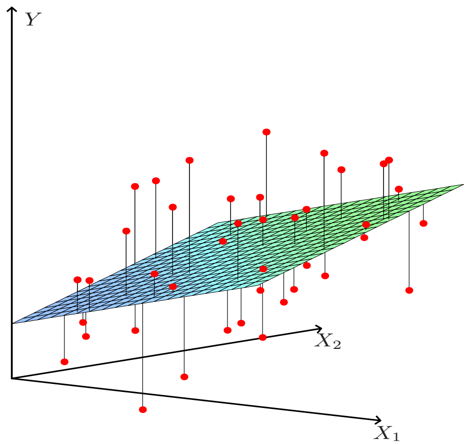
$$y_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_m x_{i,m} \quad (i = 1, \ldots, n)$$

where $\beta_j, j = 0, \ldots, m$ are the model parameters

The relationship between the input and output variables can also be expressed as follows:

$$\boldsymbol{y} = \begin{bmatrix} 16{,}000 \\ \vdots \\ 13{,}500 \end{bmatrix} = \begin{bmatrix} 1 & 25000 & 3 & 6 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 45{,}000 & 3 & 12 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_m \end{bmatrix} = \boldsymbol{X\beta}$$

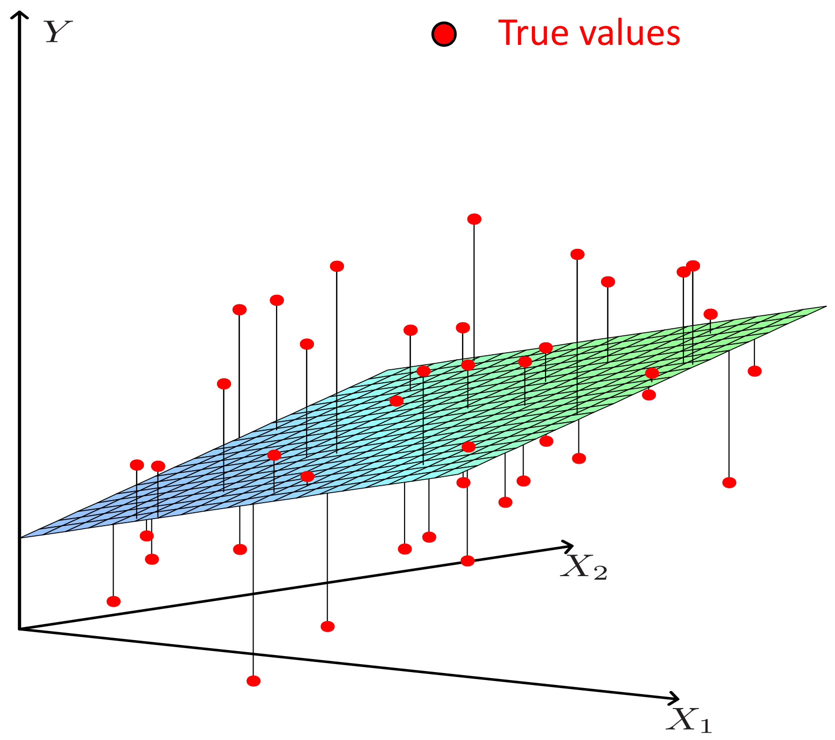
Geometrical Interpretation



Input Variable (Mileage, Km)	*# doors*	*# cylinders*	*Output Variable (Cost, £)*
25,000	3	6	16,000
105,000	5	12	11,500
120,000	5	8	6,000
140,000	3	8	3,000
45,000	3	12	13,500

Multivariate 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

Note that true response values and the predicted response values are given by:

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \text{ and } \hat{\boldsymbol{y}} = \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,m} \\ 1 & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,m} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_m \end{bmatrix} = \boldsymbol{X\beta}$$

Note also that the sum-of-square residuals is given by

$$SSR(\boldsymbol{\beta}) = tr\left((\boldsymbol{y} - \hat{\boldsymbol{y}}) \cdot (\boldsymbol{y} - \hat{\boldsymbol{y}})^t\right) = tr\left((\boldsymbol{y} - \boldsymbol{X\beta}) \cdot (\boldsymbol{y} - \boldsymbol{X\beta})^t\right)$$

Multiv. Linear Regression: Learning Algorithm

Learning Algorithm

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

$$\boldsymbol{\beta}^* = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, SSR(\boldsymbol{\beta}) = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, tr\big((\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \cdot (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^t\big)$$

This leads to the optimal model parameters given by:

$$\boldsymbol{\beta}^* = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y} = \boldsymbol{X}^\dagger\boldsymbol{y}$$

New Predictions

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

$$\hat{\boldsymbol{y}} = \boldsymbol{x}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y}$$