2. Linear Regression

Shabana K M
PhD Research Scholar
Computer Science and Engineering

IIT Palakkad

7 August 2021

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Recap

- Artificial intelligence and machine learning

Recap

- Artificial intelligence and machine learning
- Datasets for machine learning

Recap

- Artificial intelligence and machine learning
- Datasets for machine learning
  - features/attributes
  - data points/observations

Recap

- Artificial intelligence and machine learning
- Datasets for machine learning
  - features/attributes
  - data points/observations
- Types of machine learning algorithms

Recap

- Artificial intelligence and machine learning
- Datasets for machine learning
  - features/attributes
  - data points/observations
- Types of machine learning algorithms
  - supervised learning: regression, classification
  - unsupervised learning
  - reinforcement learning

Recap

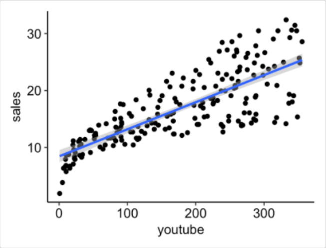
- Artificial intelligence and machine learning
- Datasets for machine learning
  - features/attributes
  - data points/observations
- Types of machine learning algorithms
  - supervised learning: regression, classification
  - unsupervised learning
  - reinforcement learning
- Supervised machine learning model

Recap

- Artificial intelligence and machine learning
- Datasets for machine learning
  - features/attributes
  - data points/observations
- Types of machine learning algorithms
  - supervised learning: regression, classification
  - unsupervised learning
  - reinforcement learning
- Supervised machine learning model
- Designing a machine learning system

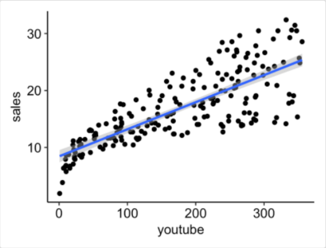
Regression: Motivation

The following scatter plot shows the impact of YouTube advertising of a company on its sales (thousands of dollars). The advertising experiment has been repeated 200 times with different budgets and the observed sales have been recorded.



Regression: Motivation

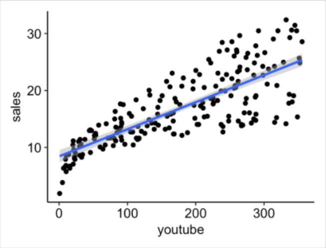
The following scatter plot shows the impact of YouTube advertising of a company on its sales (thousands of dollars). The advertising experiment has been repeated 200 times with different budgets and the observed sales have been recorded.



How can the company predict its future sales on the basis of advertising budget spent on YouTube?

Regression: Motivation

The following scatter plot shows the impact of YouTube advertising of a company on its sales (thousands of dollars). The advertising experiment has been repeated 200 times with different budgets and the observed sales have been recorded.



How can the company predict its future sales on the basis of advertising budget spent on YouTube?

Other applications: predicting the price of a house, predicting sales, weather forecasting, etc.

Regression

Regression

- method to determine the statistical relationship between a dependent variable and one or more independent variables

Regression

- method to determine the statistical relationship between a dependent variable and one or more independent variables
  - dependent variable: a quantity whose value we are trying to predict `(target variable)`

Regression

- method to determine the statistical relationship between a dependent variable and one or more independent variables
  - dependent variable: a quantity whose value we are trying to predict (`target variable`)
  - independent variables/regressors: factors that may have an impact on the dependent variable (`features`)

Regression

- method to determine the statistical relationship between a dependent variable and one or more independent variables
  - dependent variable: a quantity whose value we are trying to predict (`target variable`)
  - independent variables/regressors: factors that may have an impact on the dependent variable (`features`)
- used to predict the value of a `continuous` target variable y given a set of input features x

Regression

- method to determine the statistical relationship between a dependent variable and one or more independent variables
  - dependent variable: a quantity whose value we are trying to predict (`target variable`)
  - independent variables/regressors: factors that may have an impact on the dependent variable (`features`)
- used to predict the value of a `continuous` target variable y given a set of input features x
- supervised learning technique

Regression

- method to determine the statistical relationship between a dependent variable and one or more independent variables
  - dependent variable: a quantity whose value we are trying to predict (`target variable`)
  - independent variables/regressors: factors that may have an impact on the dependent variable (`features`)
- used to predict the value of a `continuous` target variable y given a set of input features x
- supervised learning technique
- fitting a function to a set of data points

Regression

- method to determine the statistical relationship between a dependent variable and one or more independent variables
  - dependent variable: a quantity whose value we are trying to predict (`target variable`)
  - independent variables/regressors: factors that may have an impact on the dependent variable (`features`)
- used to predict the value of a `continuous` target variable y given a set of input features x
- supervised learning technique
- fitting a function to a set of data points
- example: predicting the marks scored by students based on number of hours studied, attendance percentage, etc.

Regression

- method to determine the statistical relationship between a dependent variable and one or more independent variables
  - dependent variable: a quantity whose value we are trying to predict (`target variable`)
  - independent variables/regressors: factors that may have an impact on the dependent variable (`features`)
- used to predict the value of a `continuous` target variable y given a set of input features x
- supervised learning technique
- fitting a function to a set of data points
- example: predicting the marks scored by students based on number of hours studied, attendance percentage, etc.
  - *dependent variable*:

Regression

- method to determine the statistical relationship between a dependent variable and one or more independent variables
  - dependent variable: a quantity whose value we are trying to predict (`target variable`)
  - independent variables/regressors: factors that may have an impact on the dependent variable (`features`)
- used to predict the value of a `continuous` target variable y given a set of input features x
- supervised learning technique
- fitting a function to a set of data points
- example: predicting the marks scored by students based on number of hours studied, attendance percentage, etc.
  - *dependent variable*: marks scored
  - *independent variables*:

Regression

- method to determine the statistical relationship between a dependent variable and one or more independent variables
  - dependent variable: a quantity whose value we are trying to predict (`target variable`)
  - independent variables/regressors: factors that may have an impact on the dependent variable (`features`)
- used to predict the value of a `continuous` target variable y given a set of input features x
- supervised learning technique
- fitting a function to a set of data points
- example: predicting the marks scored by students based on number of hours studied, attendance percentage, etc.
  - *dependent variable*: marks scored
  - *independent variables*: no: of hours studied, attendance percentage, etc.

Looking at 'regression'

Given input x and output y, the task is to learn the mapping from x to y

Looking at 'regression'

Given input x and output y, the task is to learn the mapping from x to y

Let's look at some examples:

Looking at 'regression'

Given input x and output y, the task is to learn the mapping from x to y

Let's look at some examples:

x	y
2	4
5	10
6	12

When $x = 3$, $y =$

Looking at 'regression'

Given input x and output y, the task is to learn the mapping from x to y

Let's look at some examples:

x	y
2	4
5	10
6	12

When $x = 3$, $y = \mathbf{6}$

Looking at 'regression'

Given input x and output y, the task is to learn the mapping from x to y

Let's look at some examples:

x	y
2	4
5	10
6	12

When $x = 3$, $y = \mathbf{6}$

$y =$

Looking at 'regression'

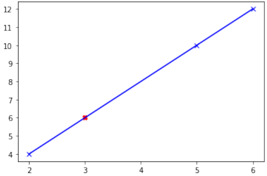
Given input x and output y, the task is to learn the mapping from x to y

Let's look at some examples:

x	y
2	4
5	10
6	12

When $x = 3$, $y = \mathbf{6}$

$$y = 2x$$



Looking at 'regression'

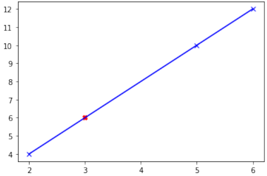
Given input x and output y, the task is to learn the mapping from x to y

Let's look at some examples:

x	y
2	4
5	10
6	12

When $x = 3$, $y = \mathbf{6}$

$y = 2x$



x	y
3	8
5	12
2	6

When $x = 6$, $y =$

Looking at 'regression'

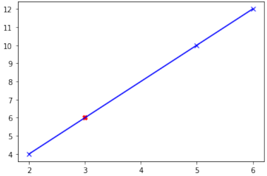
Given input x and output y, the task is to learn the mapping from x to y

Let's look at some examples:

x	y
2	4
5	10
6	12

When $x = 3$, $y = \mathbf{6}$

$y = 2x$



x	y
3	8
5	12
2	6

When $x = 6$, $y = \mathbf{14}$

Looking at ‘regression’

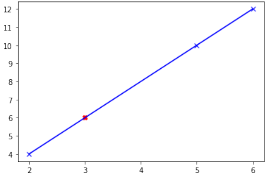
Given input x and output y, the task is to learn the mapping from x to y

Let’s look at some examples:

x	y
2	4
5	10
6	12

When $x = 3$, $y = \mathbf{6}$

$y = 2x$



x	y
3	8
5	12
2	6

When $x = 6$, $y = \mathbf{14}$

$y =$

Looking at 'regression'

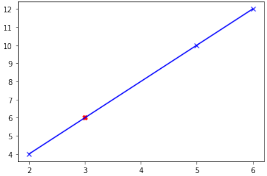
Given input x and output y, the task is to learn the mapping from x to y

Let's look at some examples:

x	y
2	4
5	10
6	12

When $x = 3$, $y = \mathbf{6}$

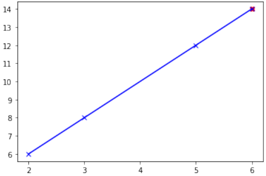
$y = 2x$



x	y
3	8
5	12
2	6

When $x = 6$, $y = \mathbf{14}$

$y = 2x + 2$



Looking at 'regression'

Let's look at some more examples:

x	y
3	18
5	50
7	98

When $x = 6$, $y =$

Looking at ‘regression’

Let’s look at some more examples:

x	y
3	18
5	50
7	98

When $x = 6$, $y =$ **72**

Looking at 'regression'

Let's look at some more examples:

x	y
3	18
5	50
7	98

When $x = 6$, $y = \mathbf{72}$

$y =$

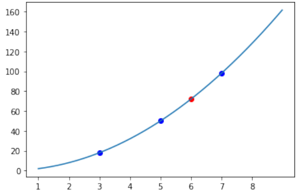
Looking at ‘regression’

Let’s look at some more examples:

x	y
3	18
5	50
7	98

When $x = 6$, $y = \mathbf{72}$

$$y = 2x^2$$



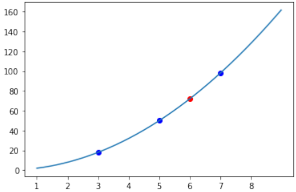
Looking at 'regression'

Let's look at some more examples:

x	y
3	18
5	50
7	98

When $x = 6$, $y = \mathbf{72}$

$$y = 2x^2$$



x	y
3	16
5	36
7	64

When $x = 6$, $y =$

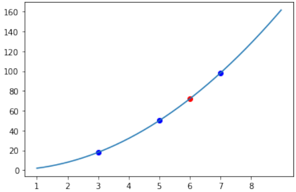
Looking at 'regression'

Let's look at some more examples:

x	y
3	18
5	50
7	98

When $x = 6$, $y = \mathbf{72}$

$$y = \; 2x^2$$



x	y
3	16
5	36
7	64

When $x = 6$, $y = \mathbf{49}$

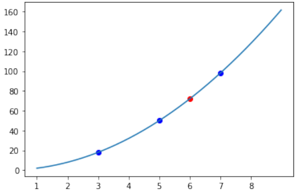
Looking at 'regression'

Let's look at some more examples:

x	y
3	18
5	50
7	98

When $x = 6$, $y = \mathbf{72}$

$y = 2x^2$



x	y
3	16
5	36
7	64

When $x = 6$, $y = \mathbf{49}$

$y =$

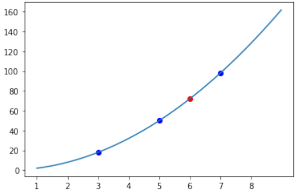
Looking at ‘regression’

Let’s look at some more examples:

x	y
3	18
5	50
7	98

When $x = 6$, $y = \mathbf{72}$

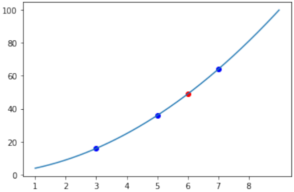
$$y = 2x^2$$



x	y
3	16
5	36
7	64

When $x = 6$, $y = \mathbf{49}$

$$y = x^2 + 2x + 1$$



Looking at ‘regression’

Let’s look at some more examples:

x_1	x_2	y
3	2	19
5	4	33
2	2	15

When $x_1 = 6, x_2 = 1,\ y =$

Looking at 'regression'

Let's look at some more examples:

x_1	x_2	y
3	2	19
5	4	33
2	2	15

When $x_1 = 6, x_2 = 1,\ y = \mathbf{28}$

Looking at 'regression'

Let's look at some more examples:

x_1	x_2	y
3	2	19
5	4	33
2	2	15

When $x_1 = 6, x_2 = 1,\ y = \mathbf{28}$

$y =$

Looking at 'regression'

Let's look at some more examples:

x_1	x_2	y
3	2	19
5	4	33
2	2	15

When $x_1 = 6, x_2 = 1,\ y = \mathbf{28}$

$$y = 4x_1 + 3x_2 + 1$$

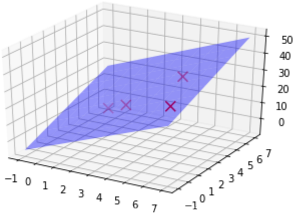
Looking at ‘regression’

Let’s look at some more examples:

x_1	x_2	y
3	2	19
5	4	33
2	2	15

When $x_1 = 6, x_2 = 1,\ y = \mathbf{28}$

$$y = 4x_1 + 3x_2 + 1$$



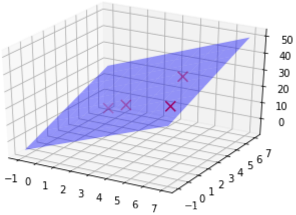
Looking at ‘regression’

Let’s look at some more examples:

x_1	x_2	y
3	2	19
5	4	33
2	2	15

When $x_1 = 6, x_2 = 1,\ y = \mathbf{28}$

$$y = 4x_1 + 3x_2 + 1$$



x	y
3	4.7
5	11.5
6	13.6
8	17.3

$y_i =$

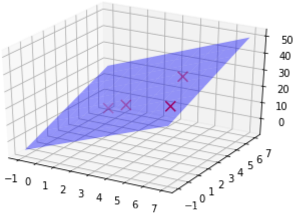
Looking at 'regression'

Let's look at some more examples:

x_1	x_2	y
3	2	19
5	4	33
2	2	15

When $x_1 = 6, x_2 = 1,\ y = \mathbf{28}$

$$y = \ 4x_1 + 3x_2 + 1$$



x	y
3	4.7
5	11.5
6	13.6
8	17.3

$$y_i = \ 2x_i + \epsilon_i, \ \ \epsilon_i \in \mathcal{R}$$

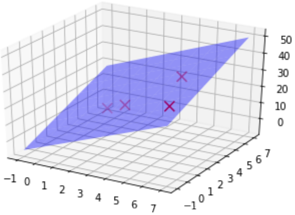
Looking at 'regression'

Let's look at some more examples:

x_1	x_2	y
3	2	19
5	4	33
2	2	15

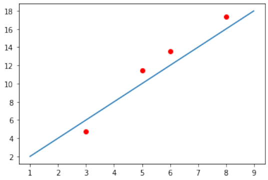
When $x_1 = 6, x_2 = 1,\ y = \mathbf{28}$

$$y = 4x_1 + 3x_2 + 1$$



x	y
3	4.7
5	11.5
6	13.6
8	17.3

$$y_i = 2x_i + \epsilon_i, \quad \epsilon_i \in \mathcal{R}$$



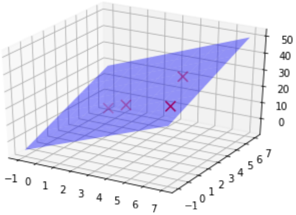
Looking at ‘regression’

Let’s look at some more examples:

x_1	x_2	y
3	2	19
5	4	33
2	2	15

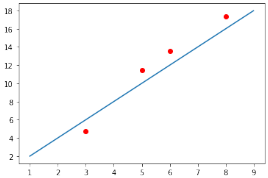
When $x_1 = 6, x_2 = 1,\ y = \mathbf{28}$

$$y = 4x_1 + 3x_2 + 1$$



x	y
3	4.7
5	11.5
6	13.6
8	17.3

$$y_i = 2x_i + \epsilon_i, \quad \epsilon_i \in \mathcal{R}$$



Not an easy task!!

Regression model

Regression model

- the dependent variable Y_i is assumed to have been generated by:

$$Y_i = f(X_i, w) + e_i$$

Regression model

- the dependent variable Y_i is assumed to have been generated by:

$$Y_i = f(X_i, w) + e_i$$

X_i represents the independent variables

Regression model

- the dependent variable Y_i is assumed to have been generated by:

$$Y_i = f(X_i, w) + e_i$$

X_i represents the independent variables

w represents the unknown parameters

Regression model

- the dependent variable Y_i is assumed to have been generated by:

$$Y_i = f(X_i, w) + e_i$$

X_i represents the independent variables
w represents the unknown parameters
e_i is an error term that represents

Regression model

- the dependent variable Y_i is assumed to have been generated by:

$$Y_i = f(X_i, w) + e_i$$

X_i represents the independent variables

w represents the unknown parameters

e_i is an error term that represents

  - the un-modelled determinants of Y_i

Regression model

- the dependent variable Y_i is assumed to have been generated by:

$$Y_i = f(X_i, w) + e_i$$

  X_i represents the independent variables
  w represents the unknown parameters
  e_i is an error term that represents
    - the un-modelled determinants of Y_i
    - random statistical noise

Regression model

- the dependent variable Y_i is assumed to have been generated by:

$$Y_i = f(X_i, w) + e_i$$

  X_i represents the independent variables

  w represents the unknown parameters

  e_i is an error term that represents

  - the un-modelled determinants of Y_i
  - random statistical noise

- goal is to estimate the function $f(X_i, w)$ that "best fits" the data

Regression model

- the dependent variable Y_i is assumed to have been generated by:

$$Y_i = f(X_i, w) + e_i$$

  X_i represents the independent variables
  w represents the unknown parameters
  e_i is an error term that represents
    - the un-modelled determinants of Y_i
    - random statistical noise
- goal is to estimate the function $f(X_i, w)$ that "best fits" the data
- for this purpose, the form of f has to be first chosen

Regression model

- the dependent variable Y_i is assumed to have been generated by:

$$Y_i = f(X_i, w) + e_i$$

  X_i represents the independent variables

  w represents the unknown parameters

  e_i is an error term that represents
    - the un-modelled determinants of Y_i
    - random statistical noise
- goal is to estimate the function $f(X_i, w)$ that "best fits" the data
- for this purpose, the form of f has to be first chosen
  - different regression models based on various forms of f

Regression model

- the dependent variable Y_i is assumed to have been generated by:

$$Y_i = f(X_i, w) + e_i$$

  X_i represents the independent variables
  w represents the unknown parameters
  e_i is an error term that represents
    - the un-modelled determinants of Y_i
    - random statistical noise
- goal is to estimate the function $f(X_i, w)$ that "best fits" the data
- for this purpose, the form of f has to be first chosen
    - different regression models based on various forms of f
- once the model is fixed, the parameters w are then estimated from the data points

Regression model

- the dependent variable Y_i is assumed to have been generated by:

$$Y_i = f(X_i, w) + e_i$$

  X_i represents the independent variables

  w represents the unknown parameters

  e_i is an error term that represents
    - the un-modelled determinants of Y_i
    - random statistical noise
- goal is to estimate the function $f(X_i, w)$ that "best fits" the data
- for this purpose, the form of f has to be first chosen
  - different regression models based on various forms of f
- once the model is fixed, the parameters w are then estimated from the data points
  - there must be sufficient data

Regression model

- the dependent variable Y_i is assumed to have been generated by:

$$Y_i = f(X_i, w) + e_i$$

  X_i represents the independent variables

  w represents the unknown parameters

  e_i is an error term that represents
    - the un-modelled determinants of Y_i
    - random statistical noise

- goal is to estimate the function $f(X_i, w)$ that "best fits" the data
- for this purpose, the form of f has to be first chosen
  - different regression models based on various forms of f
- once the model is fixed, the parameters w are then estimated from the data points
  - there must be sufficient data
  - the data points are considered to be representative of the population at large

Loss function for regression

A loss function is a measure of how good a model does in terms of predicting the expected outcome

Loss function for regression

A loss function is a measure of how good a model does in terms of predicting the expected outcome

Some of the popular regression loss functions are:

Loss function for regression

A loss function is a measure of how good a model does in terms of predicting the expected outcome

Some of the popular regression loss functions are:

1. **Mean square error/L2 loss**
   - sum of squared distances between the target and predicted values

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

Loss function for regression

A loss function is a measure of how good a model does in terms of predicting the expected outcome

Some of the popular regression loss functions are:

1. **Mean square error/L2 loss**
   - sum of squared distances between the target and predicted values

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

2. **Mean absolute error/L1 loss**
   - sum of absolute differences between the target and predicted variables

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n}$$

Loss function for regression

A loss function is a measure of how good a model does in terms of predicting the expected outcome

Some of the popular regression loss functions are:

1. **Mean square error/L2 loss**
   - sum of squared distances between the target and predicted values

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

2. **Mean absolute error/L1 loss**
   - sum of absolute differences between the target and predicted variables

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n}$$

Regression techniques estimate the weights of the model such that the value of the chosen loss function is minimized

Different regression models

Simple regression: one independent variable

Different regression models

Simple regression: one independent variable

Multiple regression: more than one independent variable

Different regression models

Simple regression: one independent variable

Multiple regression: more than one independent variable

Linear regression: assumes that the the relationship between the dependant (y) and independent (x) variables is linear

Different regression models

Simple regression: one independent variable

Multiple regression: more than one independent variable

Linear regression: assumes that the the relationship between the dependant (y) and independent (x) variables is linear

- *Simple linear regression:* $y = w_0 + w_1 x$

Different regression models

Simple regression: one independent variable

Multiple regression: more than one independent variable

Linear regression: assumes that the the relationship between the dependant (y) and independent (x) variables is linear

- *Simple linear regression:* $y = w_0 + w_1 x$
- *Multiple linear regression:* $y = w_0 + w_1 x_1 + \ldots + w_n x_n$

Different regression models

Simple regression: one independent variable

Multiple regression: more than one independent variable

Linear regression: assumes that the the relationship between the dependant (y) and independent (x) variables is linear

- *Simple linear regression:* $y = w_0 + w_1 x$
- *Multiple linear regression:* $y = w_0 + w_1 x_1 + ... + w_n x_n$

Polynomial regression: relationship between x and y modeled as an n^{th} degree polynomial in x

$$y = w_0 + w_1 x + w_2 x^2 + ... + w_n x^n$$

Different regression models

Simple regression: one independent variable

Multiple regression: more than one independent variable

Linear regression: assumes that the the relationship between the dependant (y) and independent (x) variables is linear

- *Simple linear regression:* $y = w_0 + w_1 x$
- *Multiple linear regression:* $y = w_0 + w_1 x_1 + \ldots + w_n x_n$

Polynomial regression: relationship between x and y modeled as an n^{th} degree polynomial in x

$$y = w_0 + w_1 x + w_2 x^2 + \ldots + w_n x^n$$

Non-linear regression: the relationship between x and y is modelled as $y = f(x, w)$ where f is a non-linear function of w

Different regression models

Simple regression: one independent variable

Multiple regression: more than one independent variable

Linear regression: assumes that the the relationship between the dependant (y) and independent (x) variables is linear

- *Simple linear regression:* $y = w_0 + w_1 x$
- *Multiple linear regression:* $y = w_0 + w_1 x_1 + \ldots + w_n x_n$

Polynomial regression: relationship between x and y modeled as an n^{th} degree polynomial in x

$$y = w_0 + w_1 x + w_2 x^2 + \ldots + w_n x^n$$

Non-linear regression: the relationship between x and y is modelled as $y = f(x, w)$ where f is a non-linear function of w

Logistic regression: used when the dependent variable is binary (yes/no), (0/1), etc.

A small mathematics refresher

Scalar:

24

Vector:

[2, -6, 9]

row

or column

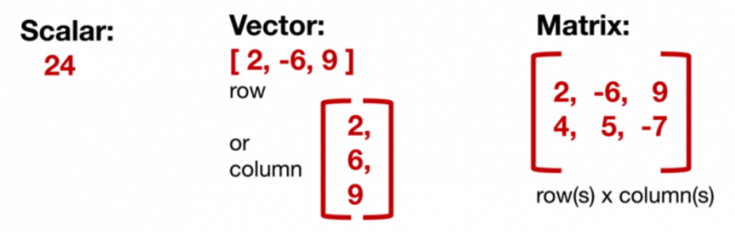
$$\begin{bmatrix} 2, \\ 6, \\ 9 \end{bmatrix}$$

Matrix:

$$\begin{bmatrix} 2, & -6, & 9 \\ 4, & 5, & -7 \end{bmatrix}$$

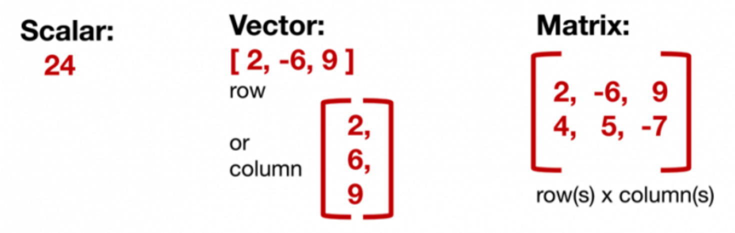
row(s) x column(s)

A small mathematics refresher



Scalar: a single number

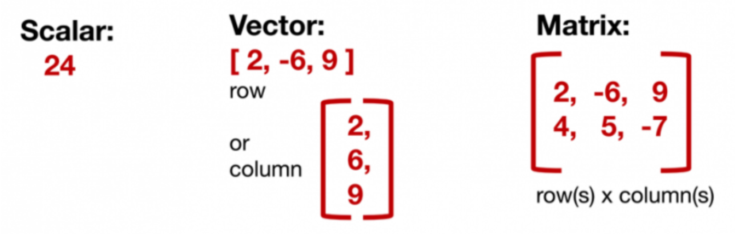
A small mathematics refresher



Scalar: a single number

Vector: an ordered array of numbers
can be in a row or a column
an index points to a specific value within the vector

A small mathematics refresher



Scalar: a single number

Vector: an ordered array of numbers
can be in a row or a column
an index points to a specific value within the vector

Matrix: two dimensional array of numbers
each element identified by two numbers

Linear regression

- assumes a linear relationship between the input variables x and target y

Linear regression

- assumes a linear relationship between the input variables x and target y
- the target variable y is modeled as a linear function of the features $x_1, x_2, \ldots, x_D$

$$\hat{y} = f(w, x) = w_0 + w_1 x_1 + \ldots + w_D x_D$$

Linear regression

- assumes a linear relationship between the input variables x and target y
- the target variable y is modeled as a linear function of the features $x_1, x_2, \dots, x_D$

$$\hat{y} = f(w, x) = w_0 + w_1 x_1 + \dots + w_D x_D$$

  - w_0 is the `intercept`

Linear regression

- assumes a linear relationship between the input variables x and target y
- the target variable y is modeled as a linear function of the features $x_1, x_2, \dots, x_D$

$$\hat{y} = f(w, x) = w_0 + w_1 x_1 + \dots + w_D x_D$$

  - w_0 is the intercept
  - $w_1, w_2, \dots, w_D$ are the parameters known as regression coefficients or weights

Linear regression

- assumes a linear relationship between the input variables x and target y
- the target variable y is modeled as a linear function of the features $x_1, x_2, ..., x_D$

$$\hat{y} = f(w, x) = w_0 + w_1 x_1 + ... + w_D x_D$$

  - w_0 is the `intercept`
  - $w_1, w_2, ..., w_D$ are the parameters known as `regression coefficients` or `weights`
- the model is a linear function of the weights and the input features

Linear regression

- assumes a linear relationship between the input variables x and target y
- the target variable y is modeled as a linear function of the features $x_1, x_2, \ldots, x_D$

$$\hat{y} = f(w, x) = w_0 + w_1 x_1 + \ldots + w_D x_D$$

  - w_0 is the `intercept`
  - $w_1, w_2, \ldots, w_D$ are the parameters known as `regression coefficients` or `weights`
- the model is a linear function of the weights and the input features
- the weights w are estimated such that the estimates of $\hat{y}$ are as close as possible to the target y values in the training set

Linear regression

- assumes a linear relationship between the input variables x and target y
- the target variable y is modeled as a linear function of the features $x_1, x_2, \ldots, x_D$

$$\hat{y} = f(w, x) = w_0 + w_1 x_1 + \ldots + w_D x_D$$

  - w_0 is the `intercept`
  - $w_1, w_2, \ldots, w_D$ are the parameters known as `regression coefficients` or `weights`
- the model is a linear function of the weights and the input features
- the weights w are estimated such that the estimates of $\hat{y}$ are as close as possible to the target y values in the training set
- the model can then be used for predicting the output y for new x values

Problem definition

Given: Training data set comprising N observations $(x_n, y_n)_{n=1}^{N}$, where $x_n = [x_{n1}, x_{n2}, ..., x_{nD}]$ is the input and y_n is the corresponding output

Problem definition

Given: Training data set comprising N observations $(x_n, y_n)_{n=1}^{N}$, where $x_n = [x_{n1}, x_{n2}, ..., x_{nD}]$ is the input and y_n is the corresponding output

Goal: Predict the y value for a new value of x

Problem definition

Given: Training data set comprising N observations $(x_n, y_n)_{n=1}^{N}$, where $x_n = [x_{n1}, x_{n2}, ..., x_{nD}]$ is the input and y_n is the corresponding output

Goal: Predict the y value for a new value of x

Estimate: The weights $w = [w_0, w_1, ..., w_D]$

Problem definition

Given: Training data set comprising N observations $(x_n, y_n)_{n=1}^{N}$, where $x_n = [x_{n1}, x_{n2}, ..., x_{nD}]$ is the input and y_n is the corresponding output

Goal: Predict the y value for a new value of x

Estimate: The weights $w = [w_0, w_1, ..., w_D]$

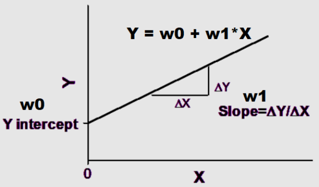
Minimize: `Mean-squared error`: $l(w) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$
where $\hat{y}_i = f(w, x_i)$

Simple linear regression

Independent Variable

Years of Experience	Salary in 1000$
2	15
3	28
5	42
13	64
8	50
16	90
11	58
1	8
9	54

Dependent Variable

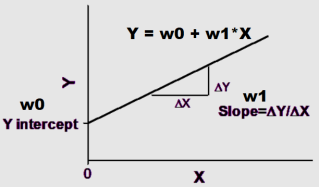


Simple linear regression

Independent Variable

Years of Experience	Salary in 1000$
2	15
3	28
5	42
13	64
8	50
16	90
11	58
1	8
9	54

Dependent Variable



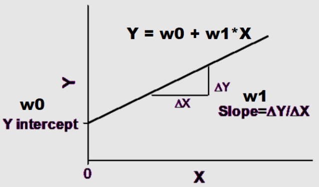
- a single independent variable x

Simple linear regression

Independent Variable

Years of Experience	Salary in 1000$
2	15
3	28
5	42
13	64
8	50
16	90
11	58
1	8
9	54

Dependent Variable



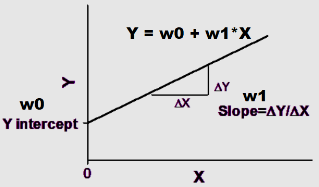
- a single independent variable x
- the model is of the form $\hat{y} = f(x, w) = w_0 + w_1 x$

Simple linear regression

Independent Variable

Years of Experience	Salary in 1000$
2	15
3	28
5	42
13	64
8	50
16	90
11	58
1	8
9	54

Dependent Variable



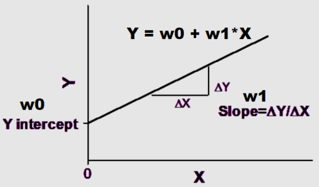
- a single independent variable x
- the model is of the form $\hat{y} = f(x, w) = w_0 + w_1 x$
- represents a line with slope w_1 and intercept w_0

Simple linear regression

Independent Variable

Years of Experience	Salary in 1000$
2	15
3	28
5	42
13	64
8	50
16	90
11	58
1	8
9	54

Dependent Variable



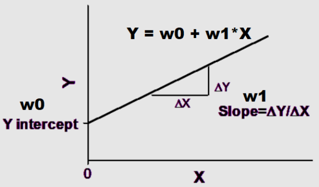
- a single independent variable x
- the model is of the form $\hat{y} = f(x, w) = w_0 + w_1 x$
- represents a line with slope w_1 and intercept w_0
- goal is to find the "best fitting" straight line to the data - `regression line`

Simple linear regression

Independent Variable

Years of Experience	Salary in 1000$
2	15
3	28
5	42
13	64
8	50
16	90
11	58
1	8
9	54

Dependent Variable



- a single independent variable x
- the model is of the form $\hat{y} = f(x, w) = w_0 + w_1 x$
- represents a line with slope w_1 and intercept w_0
- goal is to find the "best fitting" straight line to the data - `regression line`
- the weights $w = [w_0, w_1]$ are determined using an estimation technique called `Ordinary least squares`

Ordinary least squares (OLS)

- `residual` $\hat{\epsilon}_i$: difference between the actual and predicted values of the dependent variable y_i

$$\hat{\epsilon}_i = y_i - w_0 - w_1 x_i$$

Ordinary least squares (OLS)

- `residual` $\hat{\epsilon}_i$: difference between the actual and predicted values of the dependent variable y_i

$$\hat{\epsilon}_i = y_i - w_0 - w_1 x_i$$

- OLS seeks to find a line that minimizes the sum of squares of residuals, also known as `Residual Sum of Squares` (RSS) $= \sum_{i=1}^{N} \hat{\epsilon}_i^2$

Ordinary least squares (OLS)

- `residual` $\hat{\epsilon}_i$: difference between the actual and predicted values of the dependent variable y_i

$$\hat{\epsilon}_i = y_i - w_0 - w_1 x_i$$

- OLS seeks to find a line that minimizes the sum of squares of residuals, also known as `Residual Sum of Squares` (RSS) $= \sum_{i=1}^{N} \hat{\epsilon}_i^2$
- Root mean square error, (RMSE) $= \sqrt{\frac{RSS}{N}} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}}$

Ordinary least squares (OLS)

- `residual` $\hat{\epsilon}_i$: difference between the actual and predicted values of the dependent variable y_i

$$\hat{\epsilon}_i = y_i - w_0 - w_1 x_i$$

- OLS seeks to find a line that minimizes the sum of squares of residuals, also known as `Residual Sum of Squares` (RSS) $= \sum_{i=1}^{N} \hat{\epsilon}_i^2$
- Root mean square error, (RMSE) $= \sqrt{\frac{RSS}{N}} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}}$
- The regression coefficients are estimated by solving the following minimization problem:

$$\min_{w_0, w_1} l(w_0, w_1)$$

$$l(w_0, w_1) = \sum_{i=1}^{N} \hat{\epsilon}_i^2 = \sum_{i=1}^{N} (y_i - w_0 - w_1 x_i)^2$$

Ordinary least squares (OLS)

- `residual` $\hat{\epsilon}_i$: difference between the actual and predicted values of the dependent variable y_i

$$\hat{\epsilon}_i = y_i - w_0 - w_1 x_i$$

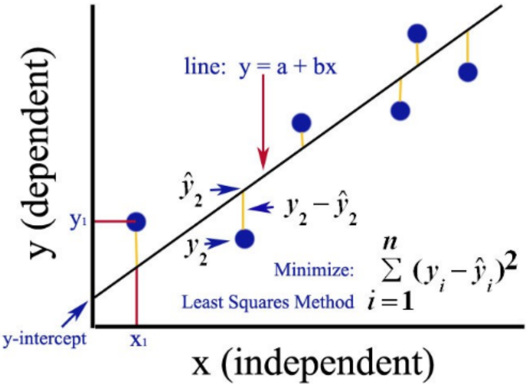
- OLS seeks to find a line that minimizes the sum of squares of residuals, also known as `Residual Sum of Squares` (RSS) $= \sum_{i=1}^{N} \hat{\epsilon}_i^2$
- Root mean square error, (RMSE) $= \sqrt{\frac{RSS}{N}} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}}$
- The regression coefficients are estimated by solving the following minimization problem:

$$\min_{w_0, w_1} l(w_0, w_1)$$

$$l(w_0, w_1) = \sum_{i=1}^{N} \hat{\epsilon}_i^2 = \sum_{i=1}^{N} (y_i - w_0 - w_1 x_i)^2$$

- can be obtained using methods of calculus

Ordinary least squares (OLS)



Ordinary least squares (OLS)

$$l(w_0, w_1) = \sum_{i=1}^{N} (y_i - w_0 - w_1 x_i)^2$$

Ordinary least squares (OLS)

$$l(w_0, w_1) = \sum_{i=1}^{N} (y_i - w_0 - w_1 x_i)^2$$

Computing w_0**:**

Ordinary least squares (OLS)

$$l(w_0, w_1) = \sum_{i=1}^{N} (y_i - w_0 - w_1 x_i)^2$$

Computing w_0**:**
Taking the partial derivative of $l(w_0, w_1)$ with respect to w_0 and equating to zero, we get:

$$w_0 = \bar{y} - w_1 \bar{x}$$

Ordinary least squares (OLS)

$$l(w_0, w_1) = \sum_{i=1}^{N} (y_i - w_0 - w_1 x_i)^2$$

Computing w_0:
Taking the partial derivative of $l(w_0, w_1)$ with respect to w_0 and equating to zero, we get:

$$w_0 = \bar{y} - w_1 \bar{x}$$

Computing w_1:

Ordinary least squares (OLS)

$$l(w_0, w_1) = \sum_{i=1}^{N} (y_i - w_0 - w_1 x_i)^2$$

Computing w_0:
Taking the partial derivative of $l(w_0, w_1)$ with respect to w_0 and equating to zero, we get:

$$w_0 = \bar{y} - w_1 \bar{x}$$

Computing w_1:
Taking the partial derivative of $l(w_0, w_1)$ with respect to w_1, substituting for w_0 from the above equation, and equating to zero, we get:

$$w_1 = \frac{N \sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N \sum_{i=1}^{N} x_i^2 - (\sum_{i=1}^{N} x_i)^2}$$

Ordinary least squares (OLS)

$$l(w_0, w_1) = \sum_{i=1}^{N} (y_i - w_0 - w_1 x_i)^2$$

Computing w_0:
Taking the partial derivative of $l(w_0, w_1)$ with respect to w_0 and equating to zero, we get:

$$w_0 = \bar{y} - w_1 \bar{x}$$

Computing w_1:
Taking the partial derivative of $l(w_0, w_1)$ with respect to w_1, substituting for w_0 from the above equation, and equating to zero, we get:

$$w_1 = \frac{N \sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N \sum_{i=1}^{N} x_i^2 - (\sum_{i=1}^{N} x_i)^2}$$

$$w_1 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N} (x_i - \bar{x})^2} = \frac{cov(x, y)}{var(x)}$$

Interpreting the regression coefficients

Applications of linear regression

Interpreting the regression coefficients

Applications of linear regression

- fit a model to predict the response for unseen observations

Interpreting the regression coefficients

Applications of linear regression

- fit a model to predict the response for unseen observations
- quantify the strength of the relationship between the response and the explanatory variables

Interpreting the regression coefficients

Applications of linear regression

- fit a model to predict the response for unseen observations
- quantify the strength of the relationship between the response and the explanatory variables

$$y = w_0 + w_1 x$$

Interpreting the regression coefficients

Applications of linear regression

- fit a model to predict the response for unseen observations
- quantify the strength of the relationship between the response and the explanatory variables

$$y = w_0 + w_1 x$$

w_0 - intercept of the model

Interpreting the regression coefficients

Applications of linear regression

- fit a model to predict the response for unseen observations
- quantify the strength of the relationship between the response and the explanatory variables

$$y = w_0 + w_1 x$$

w_0 - intercept of the model
- value taken by target variable if the dependent variable is zero

Interpreting the regression coefficients

Applications of linear regression

- fit a model to predict the response for unseen observations
- quantify the strength of the relationship between the response and the explanatory variables

$$y = w_0 + w_1 x$$

w_0 - intercept of the model
- value taken by target variable if the dependent variable is zero

w_1 - represents the rate of change of the dependent variable when the independent variable changes by one unit

Interpreting the regression coefficients

Applications of linear regression

- fit a model to predict the response for unseen observations
- quantify the strength of the relationship between the response and the explanatory variables

$$y = w_0 + w_1 x$$

w_0 - intercept of the model
- value taken by target variable if the dependent variable is zero

w_1 - represents the rate of change of the dependent variable when the independent variable changes by one unit

- $w_1 > 0$: positive relation between x and y; $x \uparrow y \uparrow$

Interpreting the regression coefficients

Applications of linear regression

- fit a model to predict the response for unseen observations
- quantify the strength of the relationship between the response and the explanatory variables

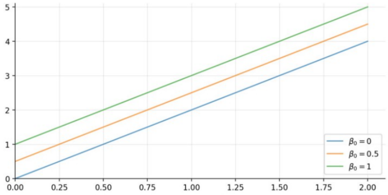
$$y = w_0 + w_1 x$$

w_0 - intercept of the model
- value taken by target variable if the dependent variable is zero

w_1 - represents the rate of change of the dependent variable when the independent variable changes by one unit

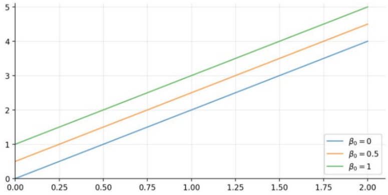
- $w_1 > 0$: positive relation between x and y; $x \uparrow y \uparrow$
- $w_1 < 0$: negative relation between x and y; $x \uparrow y \downarrow$

Interpreting the regression coefficients

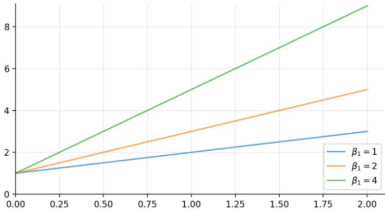


(a) Regression lines on varying the intercept value β_0 for the model: $y = \beta_0 + 2x$

Interpreting the regression coefficients



(a) Regression lines on varying the intercept value β_0 for the model: $y = \beta_0 + 2x$



(b) Regression lines on varying the coefficient value β_1 for the model: $y = 1 + \beta_1 x$

References

1. https://towardsdatascience.com/a-beginners-guide-to-regression-analysis-in-machine-learning-8a828
2. https://en.wikipedia.org/wiki/Regression_analysis
3. https://towardsdatascience.com/linear-regression-in-real-life-4a78d7159f16

Thanks Google for the pictures!