University of HUDDERSFIELD
Inspiring global professionals

Multiple Linear Regression

Lecture 17

Dr. Emmanuel Papadakis

HEA Global Teaching Excellence Award

Race Equality Charter Working Towards

Overview

- Linear relationships
- Simple Linear Regression
- Crisp functions vs Real data
- Linear Regression Model
- Training a Linear Regression Model
- Multiple Linear Regression
- Summary

Housing Price

Size in feet^2 (x)	Number of bedrooms	Age of home (years)	Price ($) in 1000's (y)
1000	3.0	12	$950,000
1750	4.0	10	$2,250,000
2200	5.0	5	$2,400,000

Linear relationships

- A statistical term used to describe a straight-line relation between two variables

Y = f(X) = m * X + b

- Y: dependent variable (target)
- X: one one more independent variables (predictors)
- f: function that predicts Y based on X
- m: coefficient of each variable = importance (slope: direction & steepness etc…)
- b: y-intercept (move up or down from (0,0)) = something like an adjustment of the value

It expresses a correlation:

how close to linear fashion **predictors change as related to the changes of the target value**

Overview

University of HUDDERSFIELD
Inspiring global professionals

- Linear relationships
- Simple Linear Regression
- Crisp functions vs Real data
- Linear Regression Model
- Training a Linear Regression Model
- Multiple Linear Regression
- Summary

Simple Linear Regression (SLR)

We want to predict Y based on the value of a variable X

- Establish if there is a relationship f so that
  - Y = **f**(X)

  - Y = b0 + b1 * X (stat world notation)
    - b0 → y-intercept (value of y when x is 0)
    - b1 → coefficient of X

Knowing relationship **f**

- We can use it to forecast new observations (regression)

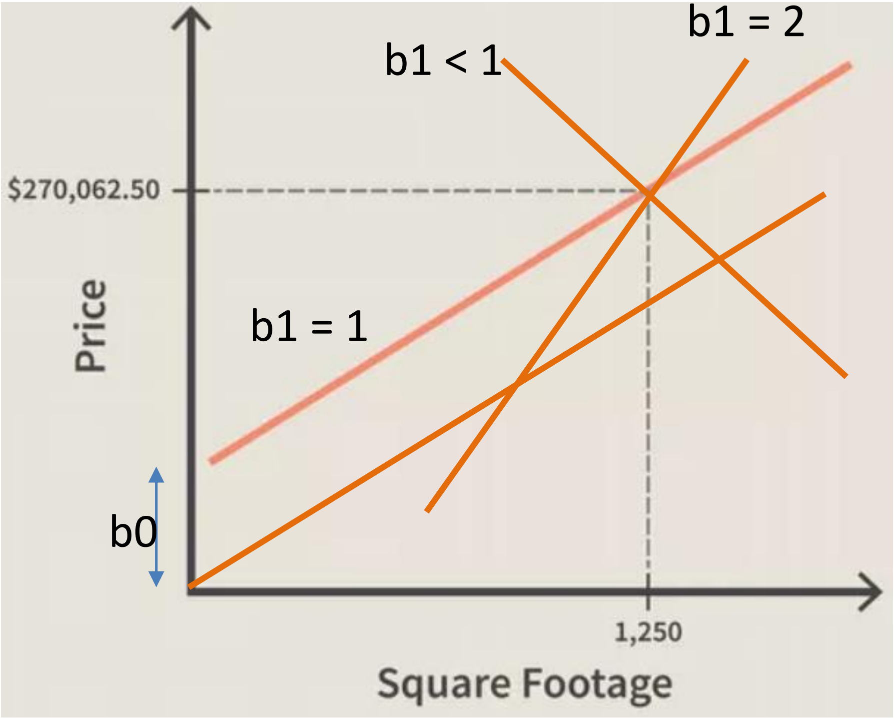
Overview

- Linear relationships
- Simple Linear Regression
- Crisp functions vs Real data
- Linear Regression Model
- Training a Linear Regression Model
- Multiple Linear Regression
- Summary

Linear relations (housing price)

- $Y = b0 + b1 * X$

Real world is not so... linear



Linear relations Vs real data

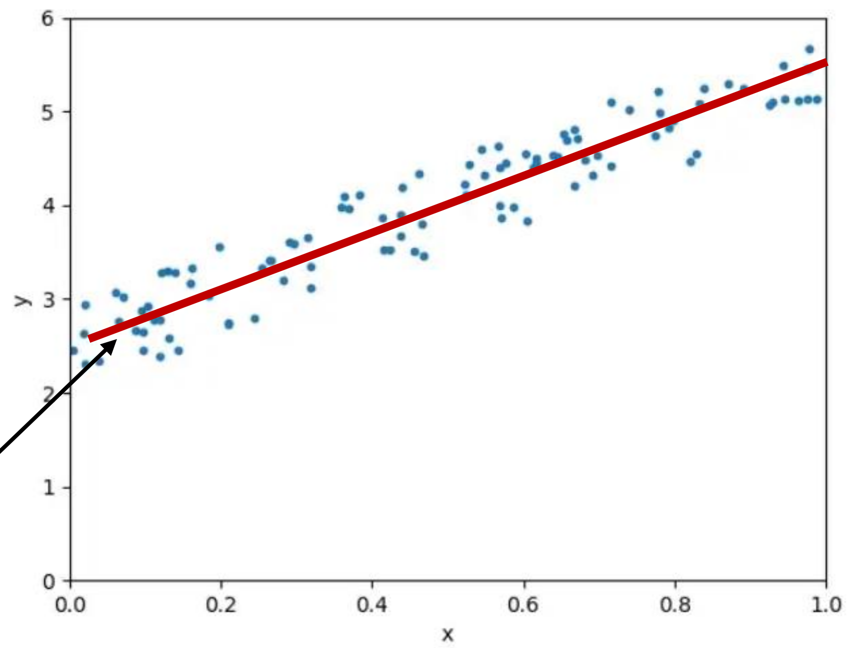
SLR: **learn** the **linear** relationship between

- a **dependent** value (**target**) and one or more **independent** variables (**predictors**)

In simple terms:

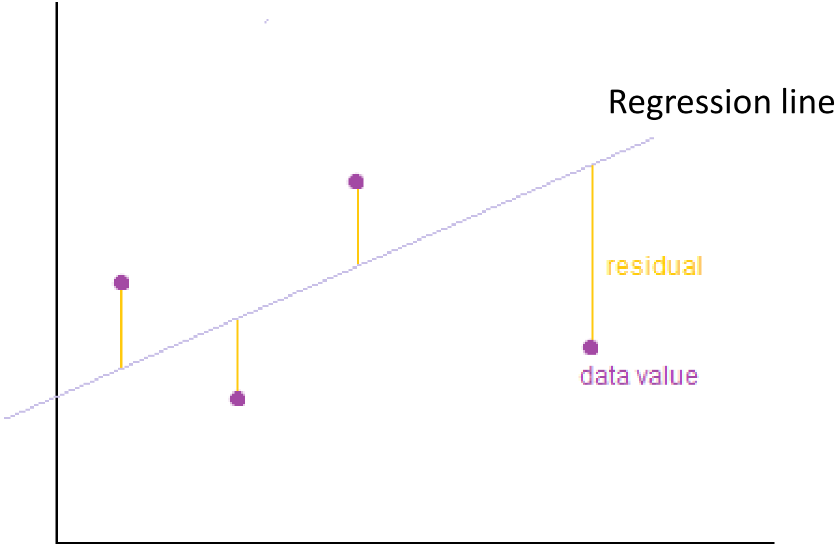
- Find the regression line that **best fits** the available data

- How?
  - Adjust y-intercepts b0 and predictors coefficients (b1) so the **regression line** represents the linear pattern of the data



Linear data patterns and residual

- Data may follow a linear pattern but the regression line **might not intersect** with any of the observations
- **Residual or error**: vertical distance between data value and regression line
- Best fit of regression line:
  - **Minimize residuals**



Overview

- Linear relationships
- Simple Linear Regression
- Crisp functions vs Real data
- Linear Regression Model
- Training a Linear Regression Model
- Multiple Linear Regression
- Summary

Linear Regression Model

Y = f(X) = b0 + b1 * X + e

- Y: dependent variable (target)
- X: independent variable (predictor)
- b0: y-intercept
- b1: coefficient of predictor X
- e: residual error

Training the model: find the best parameters b0 and b1 that minimize e

Ordinary Least Squares

Measure the error between actual and "predicted" values

Sum of the squares of the residuals

Observed value

Total number of training examples

$$MSE(B) = \frac{1}{m}\sum_{i=1}^{m}\left(f\left(x^i\right) - y^i\right)^2$$

Predicted value

A set of parameters (intercept and coefficient)

Why Squared error?

Why Squared Mean Error instead of Absolute?

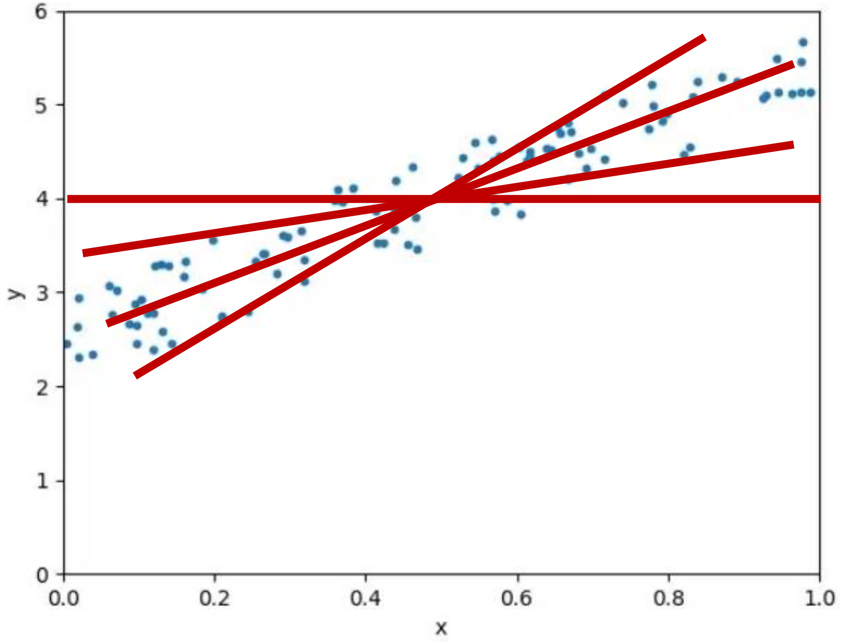
- Penalize points farther from regression
  - Distance = 2 ➔ Error = 4
  - Distance = 4 ➔ Error = 16!

Overview

University of HUDDERSFIELD
Inspiring global professionals

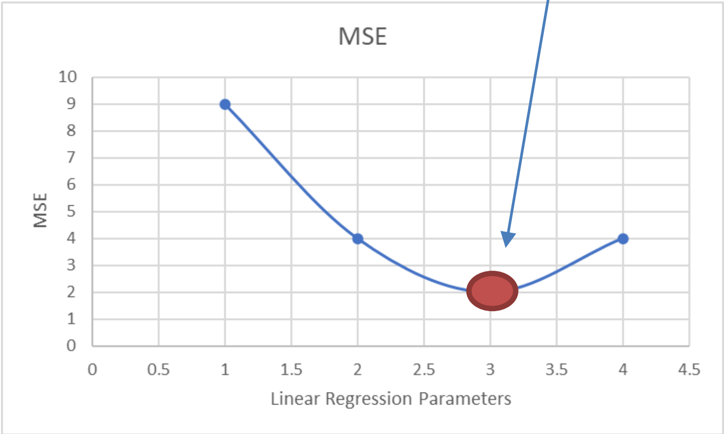
- Linear relationships
- Simple Linear Regression
- Crisp functions vs Real data
- Linear Regression Model
- Training a Linear Regression Model
- Multiple Linear Regression
- Summary

Fit line to data using OLS



B1 ➔ MSE(B1) = 9
B2 ➔ MSE(B2) = 4
B3 ➔ MSE(B3) = 2
B4 ➔ MSE(B4) = 4

Best regression line



Fit a Linear Regression Model

Gradient descent is an optimization algorithm

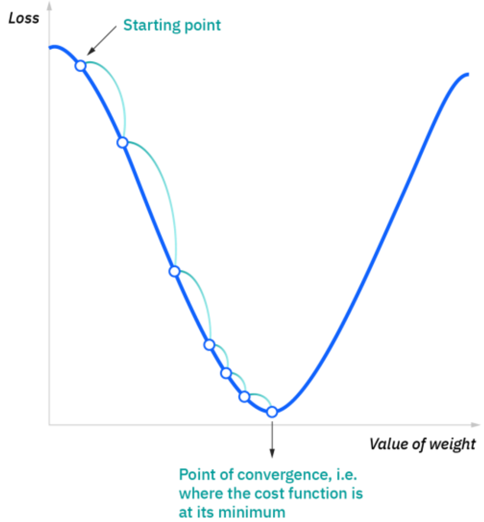
- ➢ Iteratively tweak parameters to minimize cost function
- Start with random b0 & b1
- Compute partial derivatives of MSE
  - Measure how b_j affects MSE

$$\frac{\partial SME(B)}{\partial b_j} = \frac{1}{m}\sum_{i=1}^{m}\left(f\left(x^i\right) - y^i\right)x_j^i$$

- Updated model parameters using the learning rate

$$\mathrm{b_j} = \mathrm{b_j} - \mathbf{a} * \frac{\partial SME(B)}{\partial b_j}$$

- Repeat until MSE reaches an acceptable value as low as possible

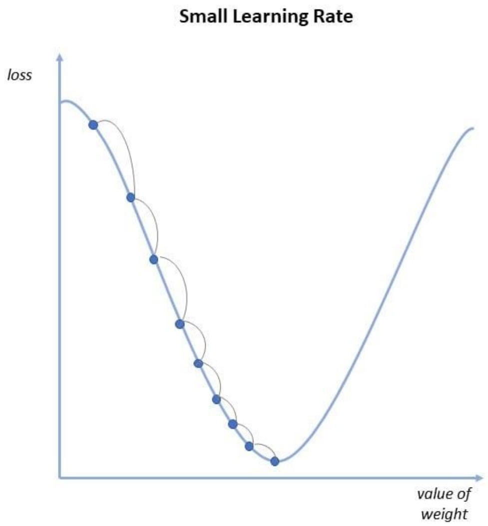


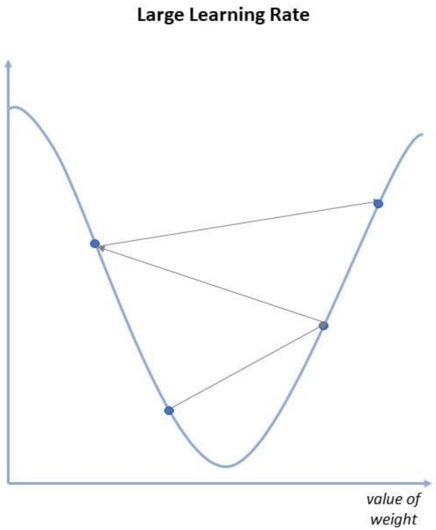
Source: ibm.com

Learning rate

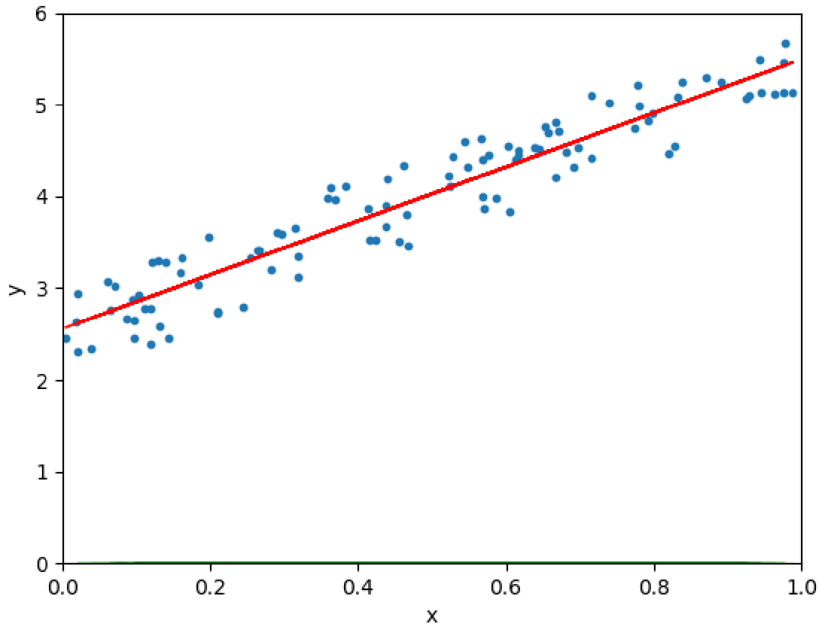
Learning rate (a) indicates how fast the cost function converges to the minimum value

- **a too small** – meticulous learning
  - Too slow & danger of overfitting!
- **a too large** – careless learning
  - Efficient but may lead to underfitting

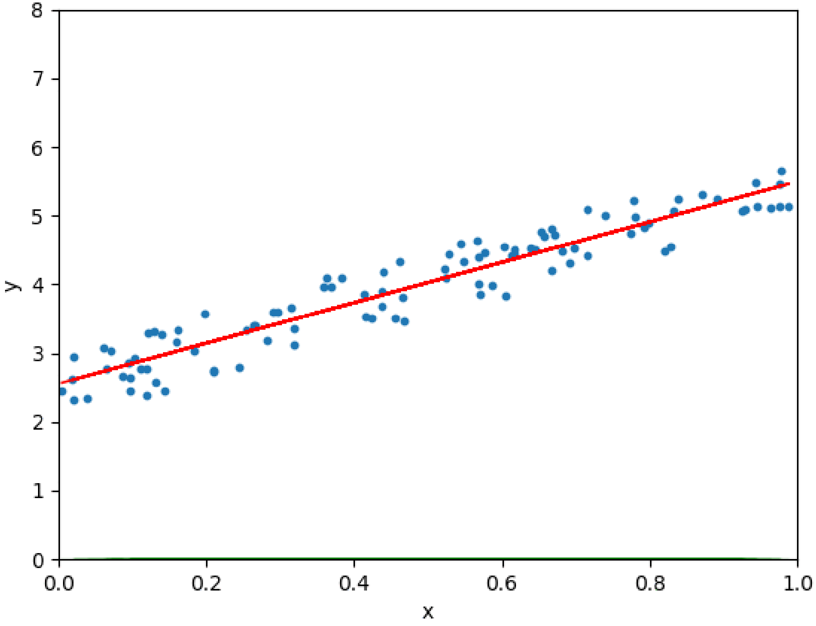




Small learning rate



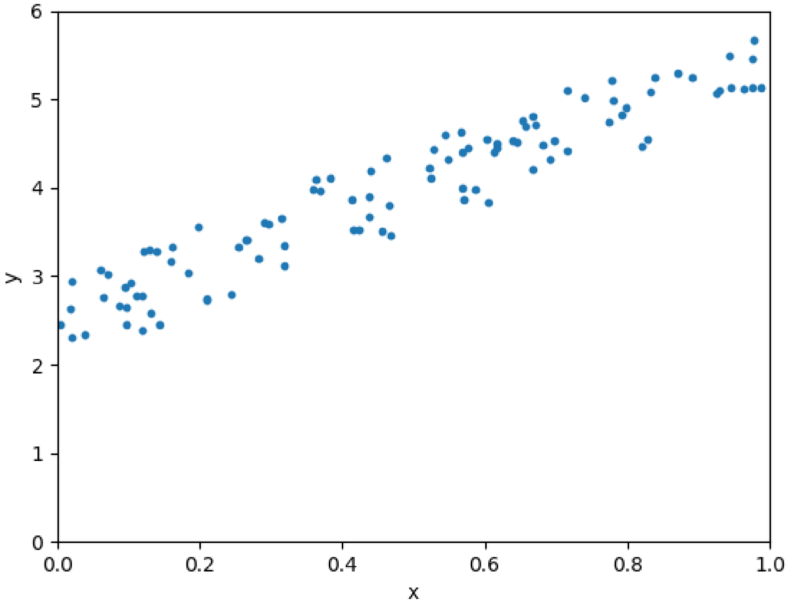
Large learning rate



Example

Assume a linear dataset

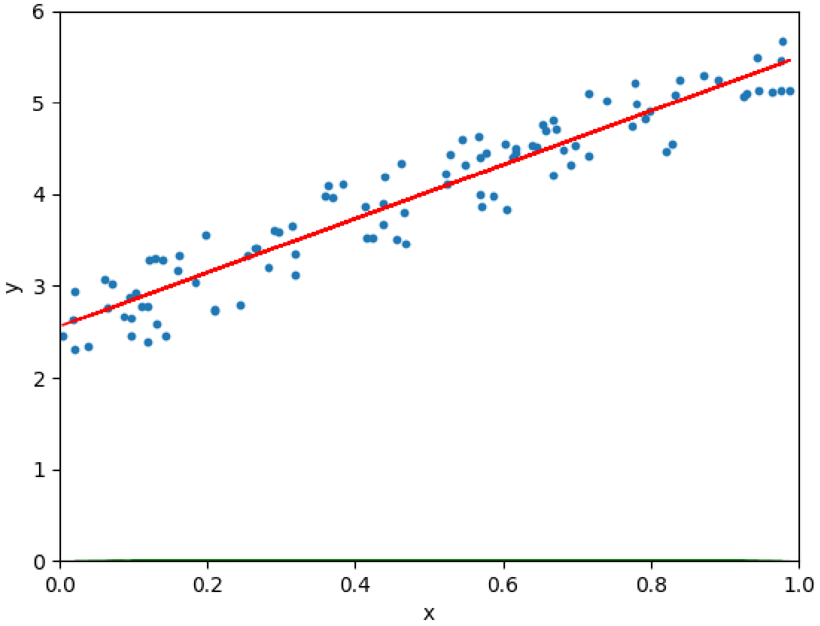
X (predictor)	Y (target)
0.548	4.32
0.715	4.54
0.602	3.51
0.544	3.9
0.42	5.11
0.64	4.82
0.43	4.41



Learning through Gradient descent

Training parameters:

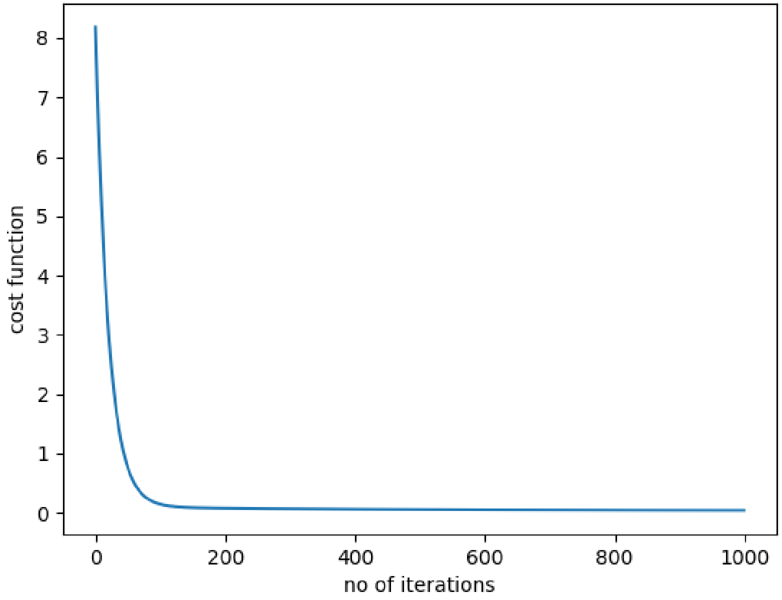
- 1000 training steps
- a = 0.01



Results of the training procedure

Plot with errors

- Best configuration when error: 0.038



Learned parameters:

- Coefficient = 2.89
- Intercept = 2.58

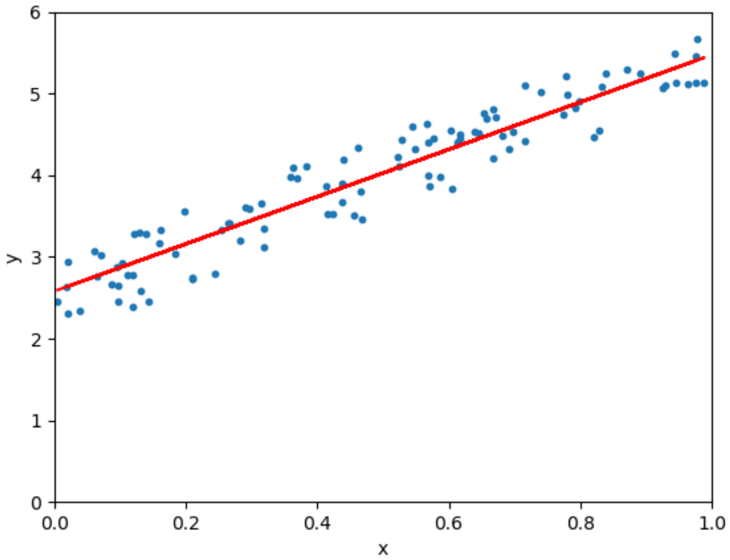
Regression line:

- Y = 2.89 * X + 2.58

Trained Linear Regression Model

The red line fits the data with the minimal error

- Forecasting new values
  - f(25) = 74.85
  - f(120) = 349.51



How well does the model perform?

Rooted Mean Squared Error: $\sqrt{MSE} = 2.76$

- Indicates the average error (lower is better)

Coefficient of determination (R^2)

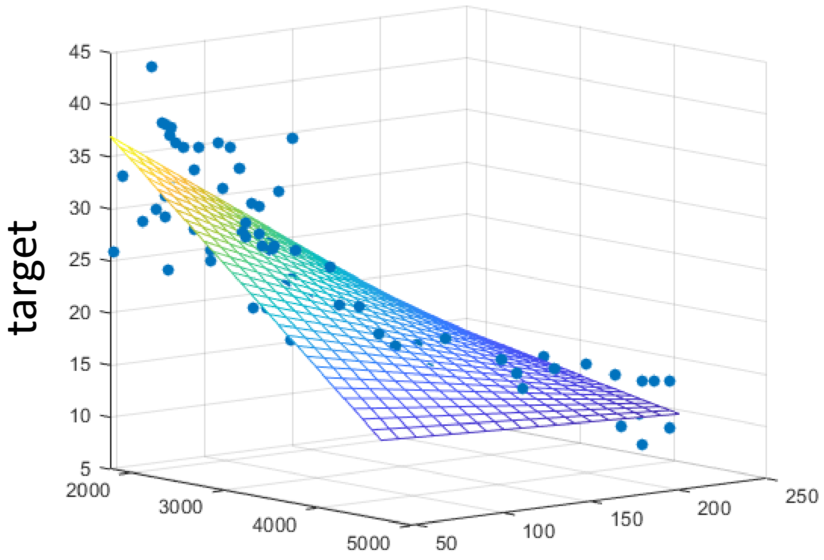
- How much the variance of Y is reduced by the MSE
  (more on that in week 12)

Overview

University of HUDDERSFIELD
Inspiring global professionals

- Linear relationships
- Simple Linear Regression
- Crisp functions vs Real data
- Linear Regression Model
- Training a Linear Regression Model
- Multiple Linear Regression
- Summary

Multiple linear regression



What happens when housing price depends on multiple criteria?

- Size, number of bedrooms and age?

Multiple Linear Regression

- An extension of linear regression in high dimensional space
- The model tries to find the best fit of n-plane over the data

$$Y = f(X) = e + b_0 + \sum_{i=1}^{n} b_n * x_n$$

- Same training routine

n^{th} variable

Coefficient of n^{th} variable

Overview

University of HUDDERSFIELD Inspiring global professionals

- Linear relationships
- Simple Linear Regression
- Crisp functions vs Real data
- Linear Regression Model
- Training a Linear Regression Model
- Multiple Linear Regression
- Summary

Main Assumptions of Linear Regression

Linear regression is easy to implement and use

- but it's predictive power is easily affected by data particularities:

Before using it, make sure that:

- Dependent and independent variables follow a linear relationship
- The order of the data do not follow any pattern
  - (e.g., error increases on larger values of x)
  - Known as autocorrelation
- There is little to no dependency between independent variables
  - known as multicollinearity
- Data observations must follow a normal distribution
  - limited number of outliers