ENRICHMENT FOR ALL

EDGE

JANUARY 8-19

Utrecht University

Regression and Demand Forecasting

Hakim Qahtan

Department of Information and Computing Sciences

Utrecht University

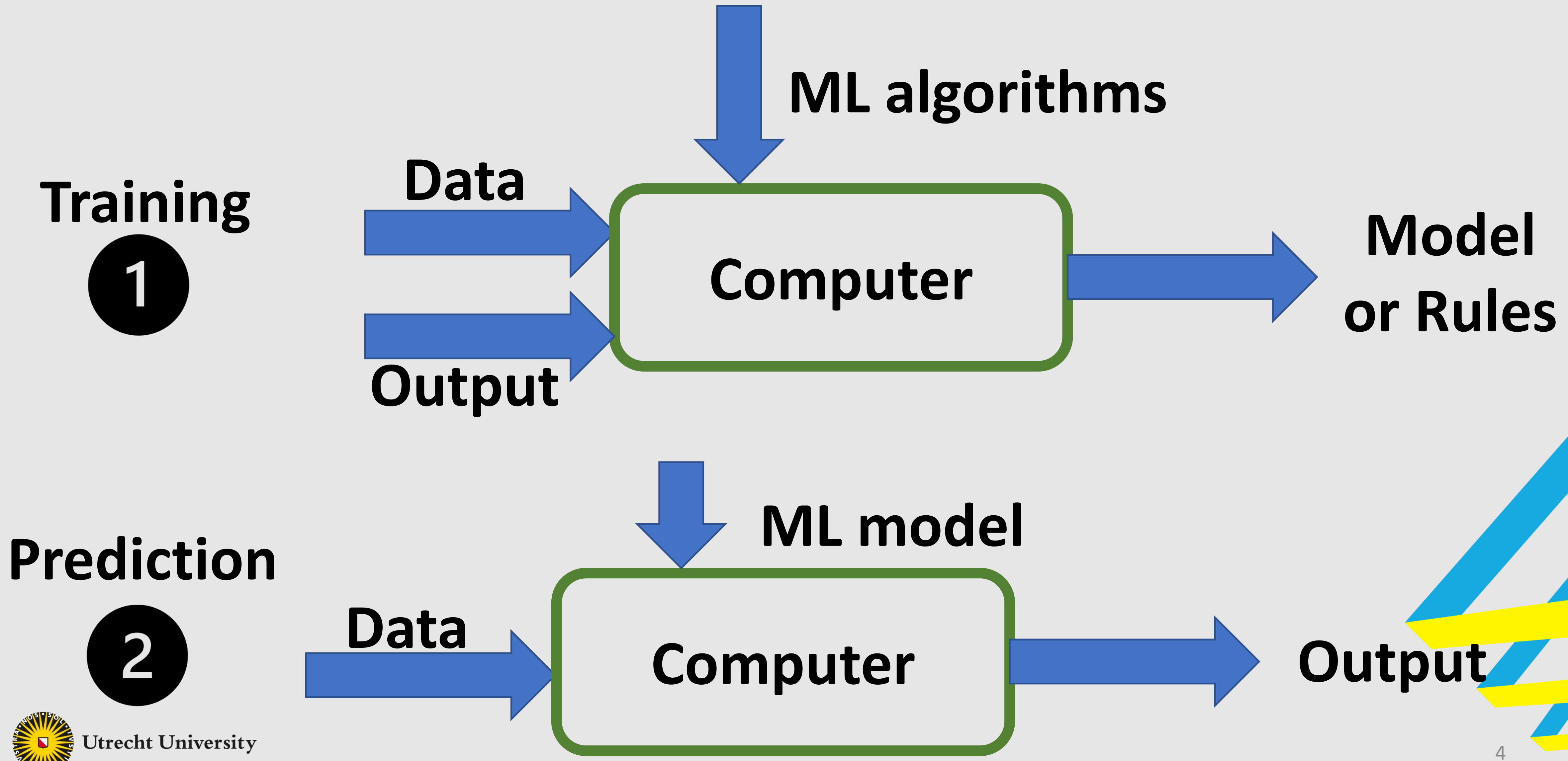
Today

Regression

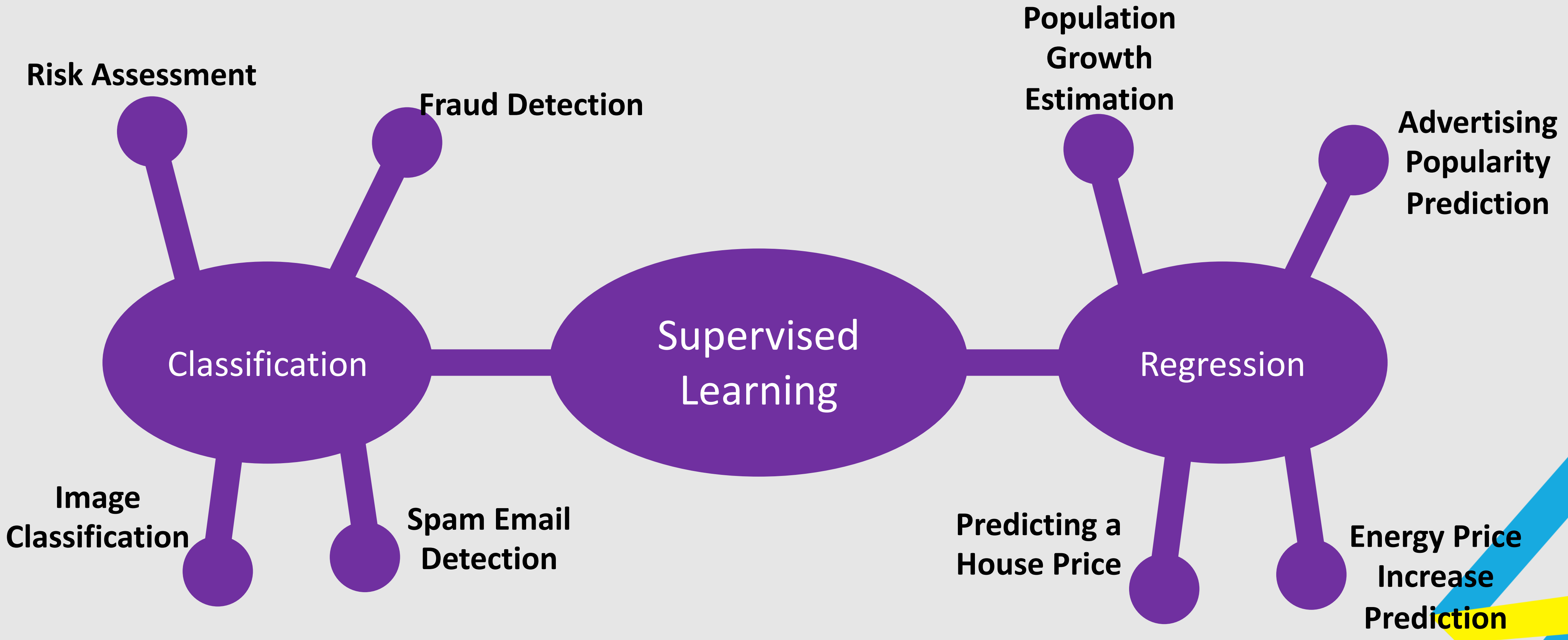
Demand Forecasting

Evaluation

Supervised Learning



Supervised Learning



Regression vs. Classification

- Regression
  - Algorithms attempt to estimate the mapping function f from the input variables x to numerical or continuous output variables y
  - Given a dataset about house prices – predict the price of a given house
- Classification
  - Algorithms attempt to estimate the mapping function f from the input variables x to discrete or categorical output variables y
  - Houses dataset – predict if the selling price is more or less than the recommended price

Regression

Regression

- Given the values of inputs X and the corresponding output Y belongs to the set of real values R, predict output accurately for new input.
- Formally:
  - Given:
    - A set of N observations $\{x_n\}_{n=1\ldots N}$ with their corresponding target values $\{y_n\}_{n=1\ldots N}$
  - Goal:
    - Predict the value of y_{n+1} for a give x_{n+1}
- Predictive technique where the target variable to be estimated is continuous

Regression (Cont.)

- Let D denote a dataset containing N observations,
$$D = \{(x_i, y_i) | i = 1,2, \dots, N\}$$
  - x_i corresponds to the values of attributes of the i-th observation.
    - These are called **explanatory variables** and can be discrete or continuous.
  - y_i corresponds to the target variable.
- Target: find a function that can minimize the error between the predicted and the actual values
  - The error can be measured as the sum of absolute or squared error

$$sum\ absolute\ error(SAE) = \sum_i |y_i - f(x_i)|$$

$$sum\ squared\ error\ (SSE) = \sum_i \left(y_i - f(x_i)\right)^2$$

What is Regression?

- Regression: making predictions about real-world quantities.
  - What would be the price of a product after producing a new version?
  - How much the discount will affect the sales volume?
  - How much the weather will affect the sales of the restaurants?
  - How many students are expected to show up in a lecture?
  - …
- Remember: regression is a supervised machine learning model

Regression – More Examples

ENRICHMENT FOR ALL

- Processes, memory → Power consumption
- Protein structure → Energy
- Heart-beat rate, age, speed, duration → Fat
- Oil supply, consumption, etc. → Oil price
- …

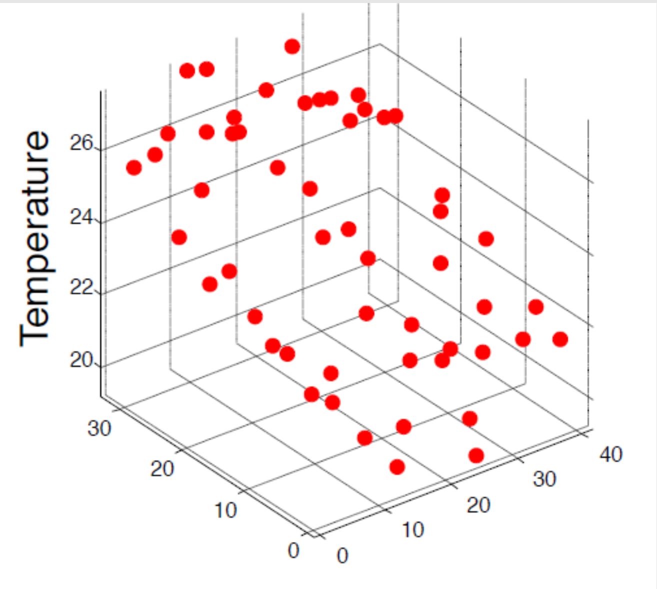
Continuous Variables

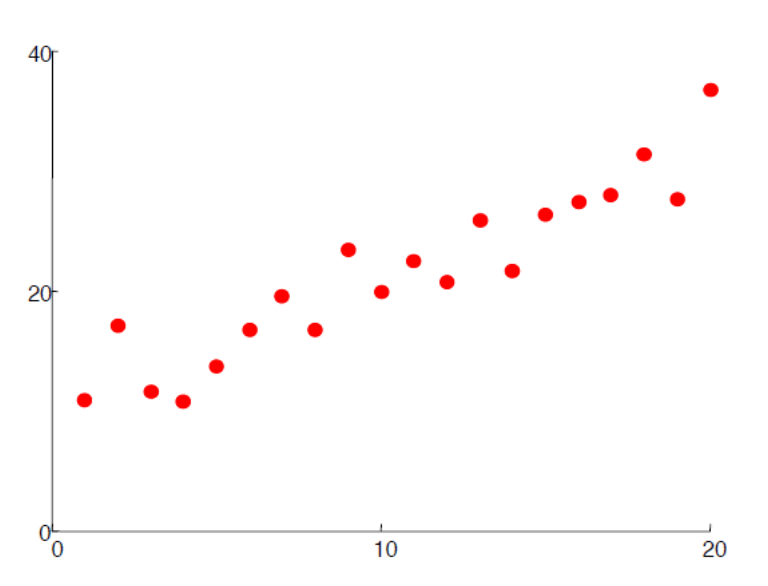
- **Definition**: Regression is the task of learning a target function f that maps each attribute set x into a continuous-valued output y.

Regression – Examples

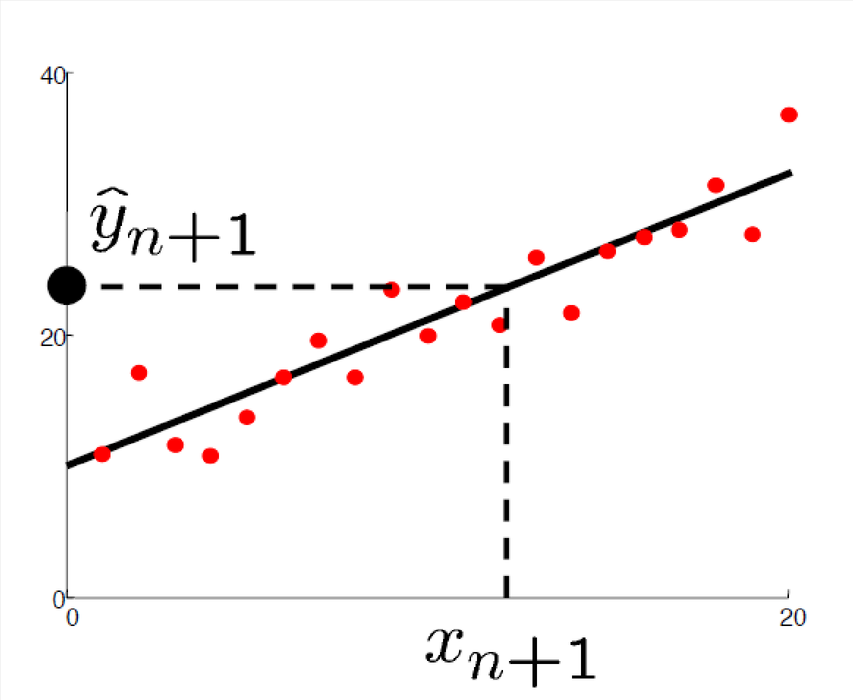
Given $\{x_n, y_n\}_{n=1 \ldots N}$

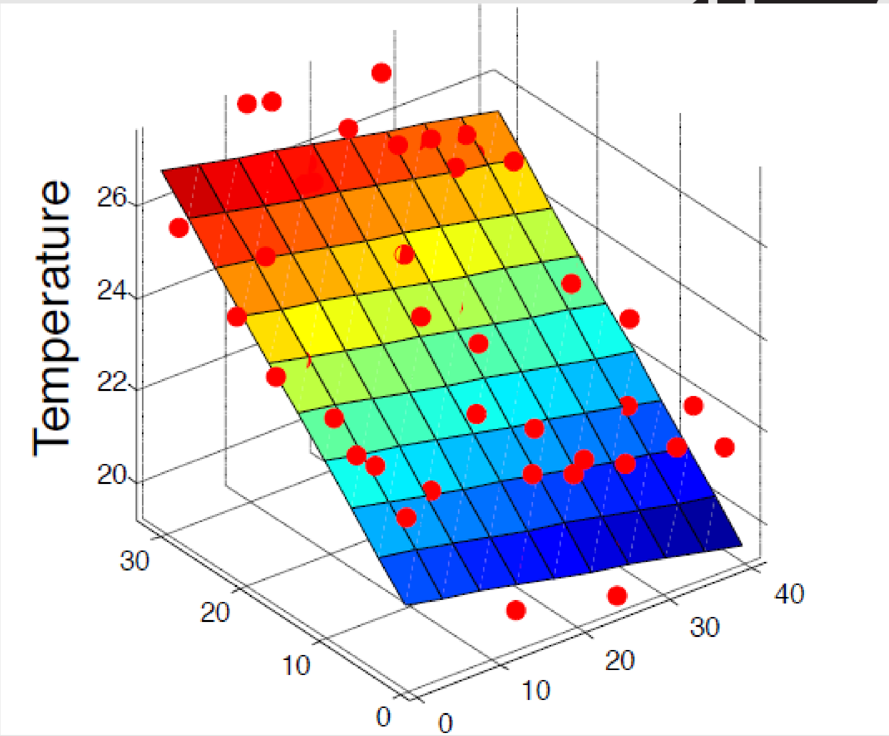
Predict the value of y_{n+1} for a give x_{n+1}





Regression – Examples



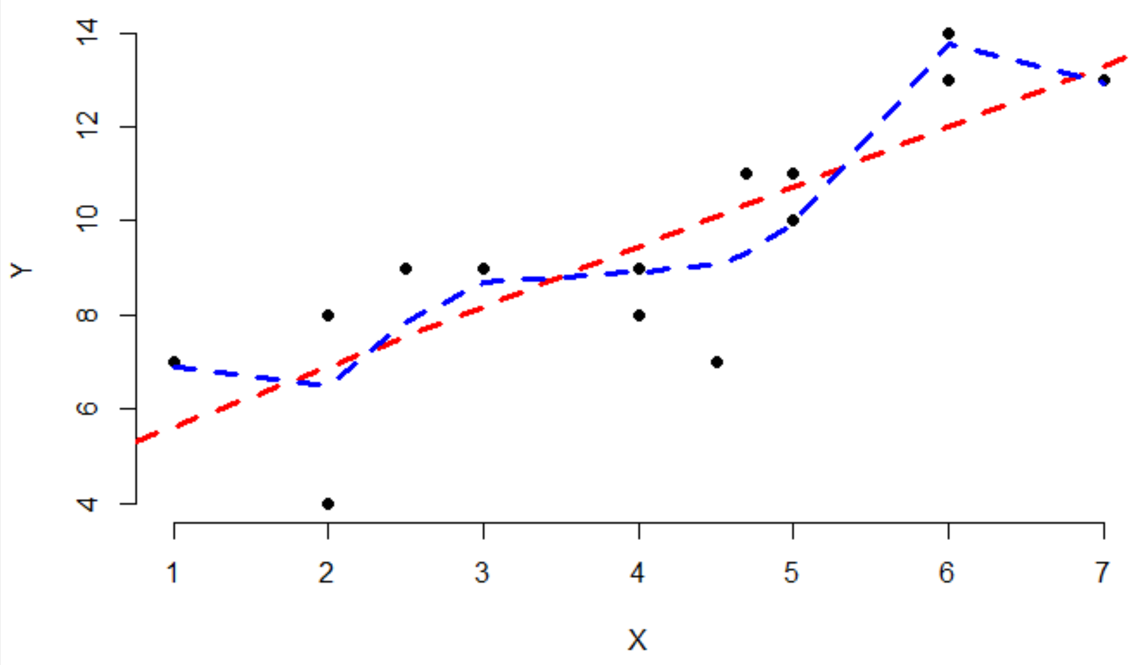


- Prediction $\hat{y}_i = w_0 + w_1 x_i$
- Prediction $\hat{y}_i = w_0 + w_1 x_{i,1} + w_2 x_{i,2}$

$$= \begin{pmatrix} 1 & x_{i,1} & x_{i,2} \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix}$$

- This is called linear regression as the function is linear in the parameters $w_{0,}\, w_1$ *and* w_2.

Simple Linear Regression



Given a set of points (x_i, y_i) such as the points in the scatterplot, find the best fitting line

$$f(x_i) = \omega_0 + \omega_1 x_i$$

such that:

$$SSE = \sum_i \big(y_i - f(x_i)\big)^2$$

$$= \sum_i (y_i - \omega_0 - \omega_1 x_i)^2$$

is minimized

Simple Linear Regression (Cont.)

- The above optimization problem can be solved by:
  1. Taking the partial derivatives of SSE with respect to ω_0 and ω_1
  2. Setting $\frac{\partial SSE}{\partial \omega_0}$ and $\frac{\partial SSE}{\partial \omega_1}$ to 0
  3. Solving the system of linear equations

     Since: $SSE = \sum_i (y_i - \omega_0 - \omega_1 x_i)^2$

     Then $\frac{\partial SSE}{\partial \omega_0} = -2\sum_i (y_i - \omega_0 - \omega_1 x_i) = 0$

     And $\frac{\partial SSE}{\partial \omega_1} = -2\sum_i x_i (y_i - \omega_0 - \omega_1 x_i) = 0$

Simple Linear Regression (Cont.)

- The equations can be summarized by the normal equation:

$$\begin{pmatrix} N & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix} \begin{pmatrix} \omega_0 \\ \omega_1 \end{pmatrix} = \begin{pmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{pmatrix}$$

Example

- Consider the following dataset

x	1	2	2.5	2	4	4	4	4.5	4.7	5	5	6	6	7
y	7	4	9	8	9	8	9	7	11	11	10	13	14	13

$$\sum_i x_i = 57.7 \qquad \sum_i x_i^2 = 276.59$$

$$\sum_i y_i = 133 \qquad \sum_i x_i y_i = 598.7$$

Example (Cont.)

x	1	2	2.5	2	4	4	4	4.5	4.7	5	5	6	6	7
y	7	4	9	8	9	8	9	7	11	11	10	13	14	13

$$\sum_i x_i = 57.7 \qquad \sum_i x_i^2 = 276.59 \qquad \sum_i y_i = 133 \qquad \sum_i x_i y_i = 598.7$$

$$\begin{pmatrix} 14 & 57.7 \\ 57.7 & 276.59 \end{pmatrix} \begin{pmatrix} \omega_0 \\ \omega_1 \end{pmatrix} = \begin{pmatrix} 133 \\ 598.7 \end{pmatrix}$$

By solving the equations, we get:

$\omega_0 \approx 4.13$ and $\omega_1 \approx 1.3$

Hence:

$$f(x_i) = 1.3\, x_i + 4.13$$

Simple Linear Regression (Cont.)

- A general solution for the normal equation can be found as follows:

$$\omega_o = \bar{y} - \omega_1 \bar{x}$$

and

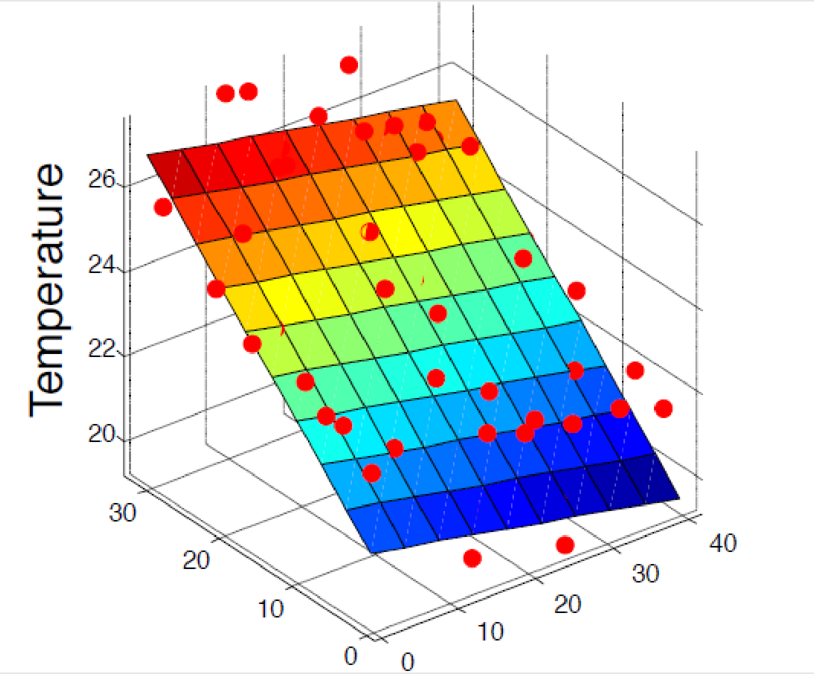
$$\omega_1 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}$$

where

$\bar{x}, \bar{y}$ are the mean (average) values for the vectors x, y

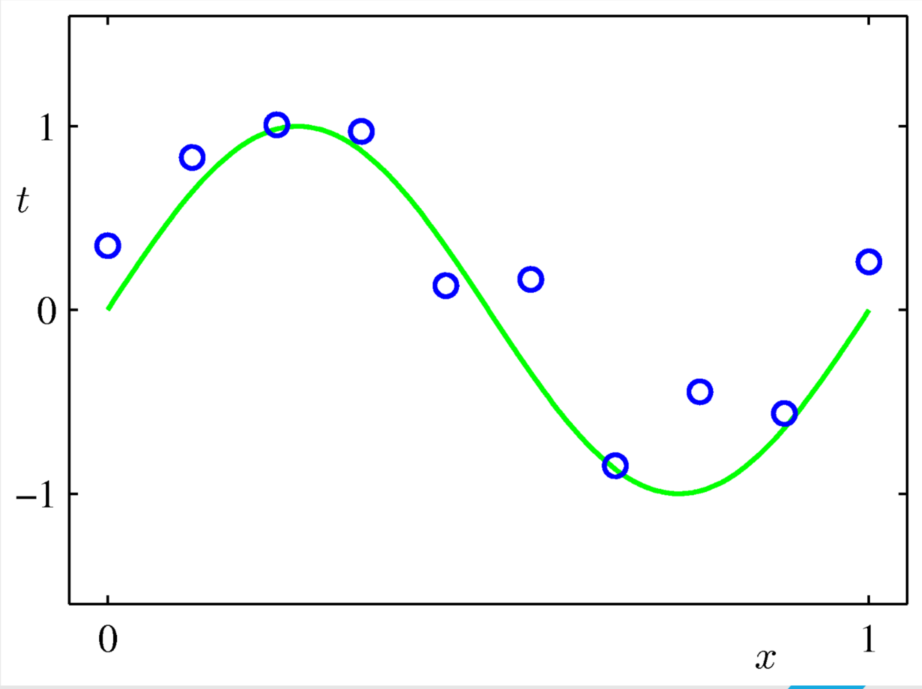
Multiple Linear Regression

- Fitting d-dimensional hyperplane to the d variables
- Simple, yet powerful
- If the relation between the response and independent variables is non-linear, we can use non-linear transformation of the variables
- E.g. x_i^2 instead of x_i



Polynomial Regression

- Suitable when the relationship between the response and the independent variables is non-linear
- Higher order polynomials complicate the model
- May cause model overfitting
- Increase the computational complexity
- Case Study: Predicting the price of a new housing market – check the provided notebook



Linear Regression in Python

```
from sklearn.linear_model import LinearRegression
import matplotlib

df = pd.read_csv('sample_data/diabetes_data.csv')
matplotlib.rcParams.update({'font.size': 18})
x = np.array(df["BMXWAIST"]).reshape(-1,1)
y = np.array(df["BMXWT"]).reshape(-1,1)
mdl  = LinearRegression()
mdl.fit(x,  y)
test  = np.arange(min(x),max(x)+4).reshape(-1,1)
test_hat  = mdl.predict(test)
```

Polynomial Regression in Python

```
x1 = np.array(df["BMXWAIST"])

y1 = np.array(df["BMXWT"])

poly_mdl = numpy.poly1d(numpy.polyfit(x1, y1, 3))

te = np.arange(min(x),max(x))

Test_hat = poly_mdl(test)
```

Degree 3

Try Larger Degrees

ENRICHMENT
FOR ALL

Coffee Break

Demand Forecasting

Decisions that Require Forecasting

- What products to produce?
- How many people to hire?
- How many units to purchase?
- How many units to produce?
- How many items to order?
- And so on……

Common Characteristics of Forecasting

- Forecasts are rarely perfect
- Forecasts are more accurate for aggregated data than for individual items
- Forecast are more accurate for shorter than longer time periods

Why Forecasting is Important?

Home > Topics > Food >

Search

Cutting down on food waste

About one-third of all the world's food goes to waste. The government wants to cut back on food waste in the Netherlands.

Huge quantities of uneaten food

A great deal of perfectly good food is never eaten and gets thrown away in household rubbish bins. Large quantities of food are also lost in harvesting, storage and transportation. Food waste is not just a waste of money. It also wastes valuable resources like water, soil and energy. The government is committed to reducing food waste and food losses.

Global Waste Facts

- There is enough food produced in the world to feed everyone. +
- One third of all food produced is lost or wasted –around 1.3 billion tonnes of food –costing the global economy close to $940 billion each year. +
- Up to 10% of global greenhouse gases comes from food that is produced, but not eaten. +
- Wasting food is worse than total emissions from flying (1.9%), plastic production (3.8%) and oil extraction (3.8%). +
- If food waste was a country, it would be the third biggest emitter of greenhouse gases after USA and China. +

Methods of Forecasting the Trend

ENRICHMENT FOR ALL

- Naïve Forecasting
- Simple Mean
- Moving Average
- Weighted Moving Average
- Exponential Smoothing
- Regression models

Methods of Forecasting the Trend – Example

- Determine forecast for periods 11
  - Naïve forecast
  - Simple average
  - 3- and 5-period moving average
  - 3-period weighted moving average with weights 0.5, 0.3, and 0.2
  - Exponential smoothing with alpha=0.2 and 0.5

Period	Orders
1	122
2	91
3	100
4	77
5	115
6	58
7	75
8	128
9	111
10	88

Methods of Forecasting the Trend – Naïve Forecasting

- Next period's forecast = previous period's actual

$$\hat{y}_{t+1} = y_t$$

$\hat{y}_t$ represents the predicted value at time t

y represents the actual value at time t

Period	Orders	Naïve Forecast
1	122	
2	91	122
3	100	91
4	77	100
5	115	77
6	58	115
7	75	58
8	128	75
9	111	128
10	88	111
11		88

Methods of Forecasting the Trend – Simple Average

- Next period's forecast = average of previously overserved data

$$\hat{y}_{t+1} = \frac{y_1 + y_2 + \cdots + y_t}{t}$$

Period	Orders	Simple Average
1	122	
2	91	122
3	100	107
4	77	104
5	115	98
6	58	101
7	75	94
8	128	91
9	111	96
10	88	97
11		97

ENRICHMENT FOR ALL

Utrecht University

Methods of Forecasting the Trend – Moving Average

- Next period's forecast = simple average of the last k periods

$$\hat{y}_{t+1} = \frac{y_{t-k+1} + y_{t-k+2} + \cdots + y_t}{k}$$

- Also called **Rolling Window**
- A smaller k makes the forecast more **responsive**
- A larger k makes the forecast more **stable**

Period	Orders	Moving Average (k = 3)	Moving Average (k = 5)
1	122		
2	91		
3	100		
4	77	104	
5	115	89	
6	58	97	101
7	75	83	88
8	128	83	85
9	111	87	91
10	88	105	97
11		109	92

Methods of Forecasting the Trend – Weighted Moving Average

- Next period's forecast = weighted average of the last k periods

$$\hat{y}_{t+1} = c_1 y_{t-k+1} + \cdots + c_k y_t$$

With

$$c_1 + c_2 + \cdots + c_k = 1$$

We take $c_1 = 0.2$, $c_2 = 0.3$ and $c_3 = 0.5$

Period	Orders	Weighted Moving Average (k = 3)
1	122	
2	91	
3	100	
4	77	102
5	115	87
6	58	101
7	75	79
8	128	78
9	111	98
10	88	109
11		103

Methods of Forecasting the Trend – Exponential Smoothing

- Next period's forecast = weighted average of the previous reading and the history

$$\hat{y}_{t+1} = \alpha\, y_t + (1-\alpha)\hat{y}_t$$

$$\hat{y}_3 = 0.2 * 91 + 0.8 * 122 = 116$$

- A smaller α makes the forecast more **stable**
- A larger α makes the forecast more **responsive**

Period	Orders	Exponential Smoothing($\alpha = 0.2$)	Exponential Smoothing($\alpha = 0.5$)
1	122		
2	91		
3	100	116	107
4	77	113	104
5	115	106	91
6	58	108	103
7	75	98	81
8	128	93	78
9	111	100	103
10	88	102	107
11		99	98

Methods of Forecasting the Trend – Regression Models

- Training dataset
  - Include the set of features (explanatory variables) and the target variable
- Train the regression model
  - Find the best estimation of the parameters
  - $f(\boldsymbol{x}) = w_0 + w_1 x_1 + \cdots + w_n x_n$ where $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$
- Predict the value of the target variable for the new incoming record

Methods of Forecasting the Trend – Regression Models

```
from sklearn.linear_model import LinearRegression
import matplotlib
df = pd.read_csv('sample_data/diabetes_data.csv')
matplotlib.rcParams.update({'font.size': 18})
x = np.array(df["BMXWAIST"]).reshape(-1,1)
y = np.array(df["BMXWT"]).reshape(-1,1)
mdl = LinearRegression()
mdl.fit(x, y)
test = np.arange(min(x),max(x)+4).reshape(-1,1)
test_hat = mdl.predict(test)
```

Coffee Break

Evaluation

Forecast Accuracy

- Tests of forecast accuracy are based on the difference between the forecast of the variables' values at time *t* and the actual value at the same time point *t*

- The closer the two to each other $\Longrightarrow$ the smaller the forecast error, i.e. better forecast

Forecast Accuracy – Mean Squared Error (MSE)

- The MSE statistic is defined as:

$$MSE = \frac{\sum_{t=T_1}^{T}(y_t - \hat{y}_t)^2}{T - T_1 + 1}$$

- T is the total number of samples in the time series
- T_1 the index of the first value to be forecast
- $\hat{y}_t$ is the predicted value at time t
- y_t is the actual value at time t
- Another popular metric: Root Mean Squared Error (RMSE) = $\sqrt{MSE}$

Forecast Accuracy – More Metrics

- The Mean Absolute Error (MAE) :

$$MAE = \frac{\sum_{t=T_1}^{T} |(y_t - \hat{y}_t)|}{T - T_1 + 1}$$

  - It is also known as Mean Absolute Deviation (MAD)

- Tracking Signal (TS)

$$TS = \frac{\sum_{t=T_1}^{T} (y_t - \hat{y}_t)}{MAE}$$

R-Square

- R-Square (R^2) is defined as:

$$R^2 = 1 - \frac{\sum_i (f(x_i) - y_i)^2}{\sum_i (y_i - \bar{y})^2}, \qquad \hat{y}_i = f(x_i)$$

- R^2 close to 1 means that the data fits well to the regression line
- $\sum_i (f(x_i) - y_i)^2$ is called the residual sum of squares
- $\sum_i (y_i - \bar{y})^2$ is called the total sum of squares.

R-Square

- When adding more explanatory variables, the value of R^2 increases so it is adjusted using the formula

$$Adjusted\ R^2 = 1 - \left(\frac{N-1}{N-d-1}\right)(1-R^2)$$

where N is the number of data points and $d+1$ is the number of parameters of the regression model

Wrap-Up

- Summerize what you learned today in 2-minutes

Thank You

ENRICHMENT FOR ALL

Utrecht University