**Seminar 2**

**Put the data files in the same folder as your Jupyter Notebook files if there are data files.**

**We learn coding by imitation. Therefore, we start by copying example codes and run them. Based on the outputs, comments, and the codes, we understand what the codes need and what the codes produce. Then we can modify the codes and apply them to new data for solving new problems.**

1. **Try data munging: Start a new Jupyter Notebook and copy the following codes one (line/part) by one, followed by press the keys Shift + Enter.**

# Data Munging

## Data loading and preprocessing with pandas

### Fast and easy data loading

import pandas as pd

import numpy as np

iris\_filename = 'datasets-uci-iris.csv'

iris = pd.read\_csv(iris\_filename, sep=',', header=None,

names= ['sepal\_length', 'sepal\_width',

'petal\_length', 'petal\_width',

'target'])

iris.head()

iris.columns

y = iris['target']

y

X = iris[['sepal\_length', 'sepal\_width']]

X

### Dealing with problematic data

import pandas as pd

fake\_dataset = pd.read\_csv('a\_loading\_example\_1.csv', sep=',')

fake\_dataset

fake\_dataset.fillna(fake\_dataset.mean(axis=0))

### Data preprocessing

iris['target'].unique()

### Data selection

import pandas as pd

dataset = pd.read\_csv('a\_selection\_example\_1.csv')

dataset

dataset = pd.read\_csv('a\_selection\_example\_1.csv', index\_col=0)

dataset

dataset['val3'][104]

dataset.loc[104, 'val3']

dataset.iloc[4, 2]

dataset[['val3', 'val2']][0:2]

1. **Try case - trading strategies - 1: Start a new Jupyter Notebook and copy the following codes one (line/part) by one, followed by press the keys Shift + Enter.**

# Trading Strategies

import warnings

warnings.simplefilter(action='ignore', category=FutureWarning)

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.preprocessing import StandardScaler

from sklearn.model\_selection import train\_test\_split

pd.options.mode.chained\_assignment = None

# Don't show the waring when writing in a copy slice, not the original data.

## Load data

dataset = pd.read\_csv('AAPL.csv')

dataset

## Create features

dataset = dataset.dropna()

dataset = dataset[['Open', 'High', 'Low', 'Close']]

dataset['H-L'] = dataset['High'] - dataset['Low']

dataset['O-C'] = dataset['Close'] - dataset['Open']

dataset['3day MA'] = dataset['Close'].shift(1).rolling(window = 3).mean()

dataset['10day MA'] = dataset['Close'].shift(1).rolling(window = 10).mean()

dataset['30day MA'] = dataset['Close'].shift(1).rolling(window = 30).mean()

dataset['Std\_dev']= dataset['Close'].rolling(5).std()

dataset['Price\_Rise'] = np.where(dataset['Close'].shift(-1) > dataset['Close'], 1, 0)

dataset = dataset.dropna()

dataset.head()