

資料分析與策略 期末報告

經濟四 陳維廷

NTPU

2020.6.11

- 疫情對股市的影響
 - 疫情爆發前股市分析
 - 疫情爆發後股市分析
- 機器學習對於股市的預測能力探討

機器學習步驟

- 1 收集資料 (Gathering data)
- 2 準備數據 (Preparing that data)
- 3 選擇模型 (Choosing a model)
- 4 訓練機器 (Training)
- 5 評估分析 (Evaluation)
- 6 調整參數 (Hyperparameter tuning)
- 7 預測推論 (Prediction)

Lasso

- 為 Linear Regression 加上 L1 penalty function
- objective function : $minimize\{SSE + \lambda \sum_{j=1}^p \|\beta_j\|\}$

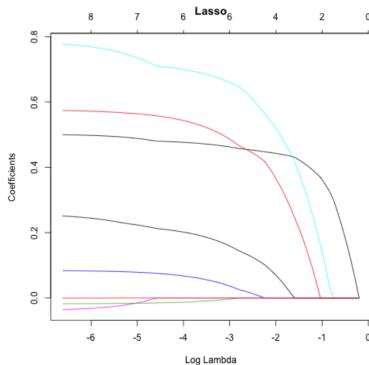


Figure: Lasso Regression

- 為Backward Selection，透過迭代逐步將重要性低的參數丟棄。
- 透過Kfold 進行cross validation。
- 與統計學Stepwise Selection的差異：評分標準不同。

Support Vector Machine



Figure: SVM Example

Support Vector Machine

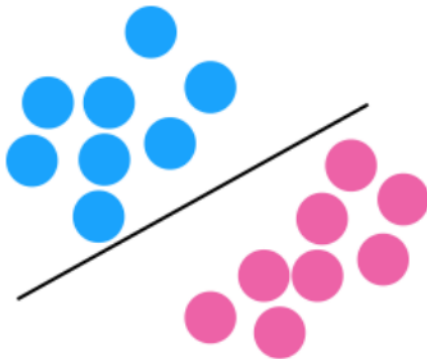


Figure: SVM Example

Support Vector Machine

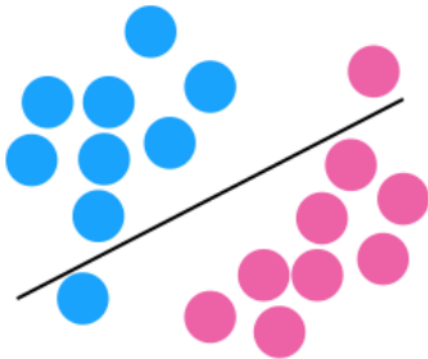


Figure: SVM Example

Support Vector Machine

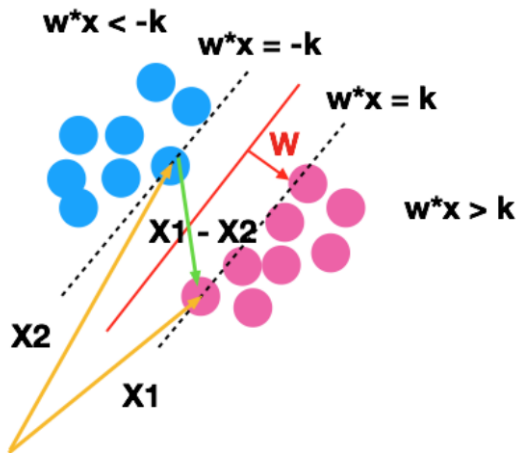


Figure: SVM Example

Support Vector Machine

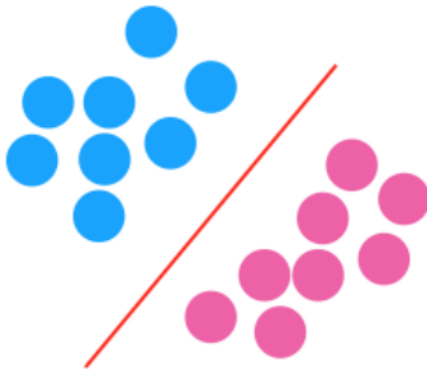


Figure: SVM Example

Support Vector Machine



Figure: SVM Example

Support Vector Machine

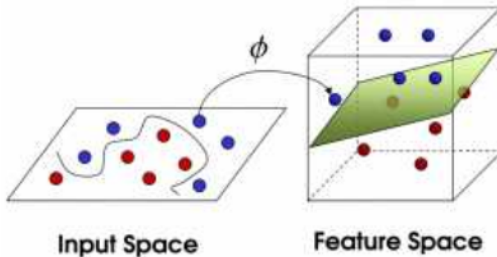


Figure: SVM Example

- objective function :

$$\left[\frac{1}{n} \sum_{i=1}^n \max\{0, 1 - y_i(\vec{w} \cdot \vec{x}_i - b)\} \right] + \textit{penalty}$$

k-Nearest Neighbors

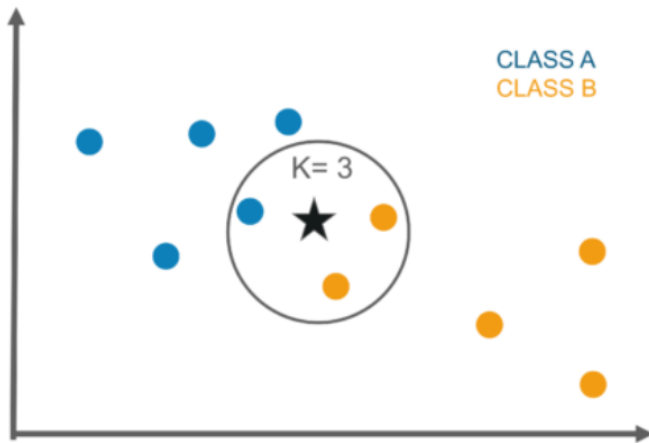


Figure: kNN

Random Forest

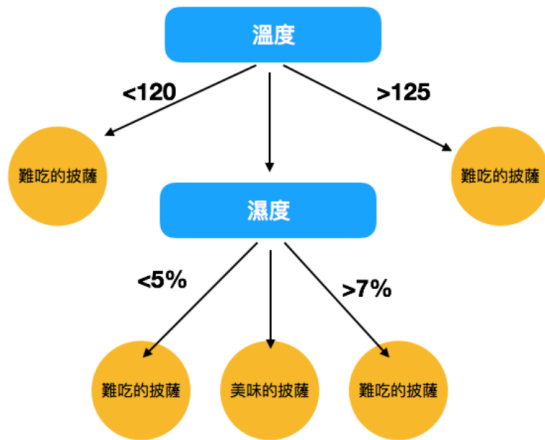


Figure: Decision Tree

Random Forest

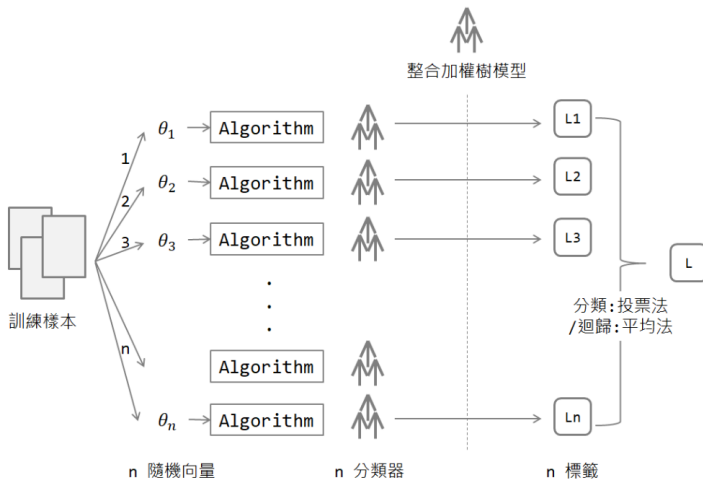


Figure: Random Forest

- 文獻：Predicting Stock Prices Using LSTM

- 步驟：

- ① Raw Data: 尋找原始資料
- ② Data Preprocessing: 資料預處理
- ③ Feature Extraction: 特徵提取
- ④ Training Neural Network: 訓練神經網路
- ⑤ Output Generation: 檢驗模型能力

Training Neural Network

```
model = Sequential()

model.add(LSTM(128, input_shape=(layers[1],
layers[0]), return_sequences=True))

model.add(LSTM(64, input_shape=(layers[1],
layers[0]), return_sequences=False))

model.add(Dense(16,init='uniform',activation='relu'))

model.add(Dense(1,init='uniform',activation='linear'))
```

Figure: Neural Network

多層感知器 (Multi-Layer Perceptron, MLP)

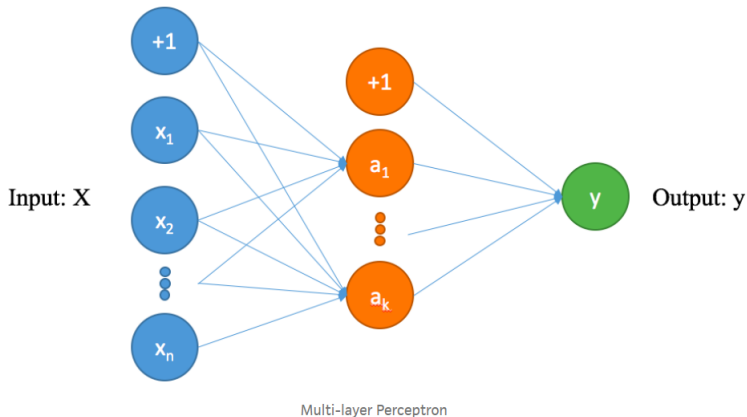


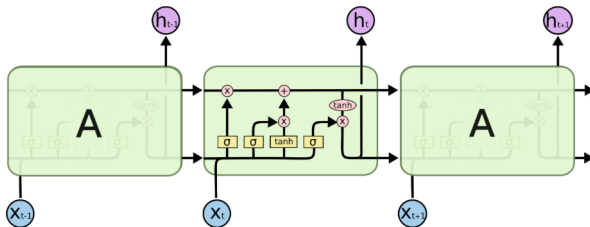
Figure: Multi-Layer Perceptron

遞歸神經網絡(Recurrent Neural Network, RNN)

- Recurrent Neural Network

長短期記憶(Long Short-Term Memory, LSTM)

- RNN的一種，而其不相同之處在於有了更多的控制單元
input gate、output gate、forget gate



The repeating module in an LSTM contains four interacting layers.

Figure: Long Short-Term Memory

Long Short-Term Memory

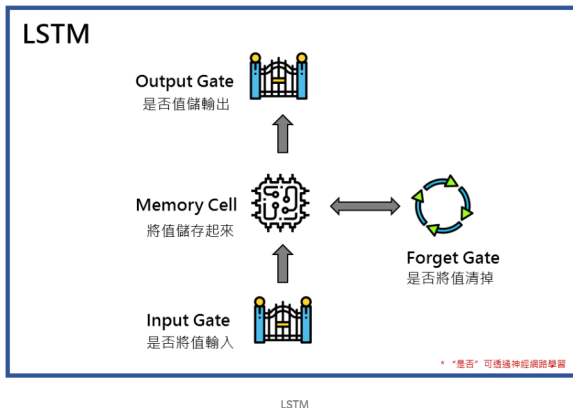
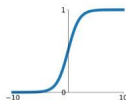


Figure: Long Short-Term Memory

Activation Functions

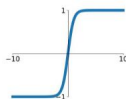
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



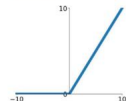
tanh

$$\tanh(x)$$



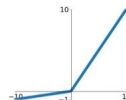
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$



Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$

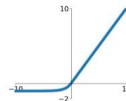


Figure: Activation Function

Output

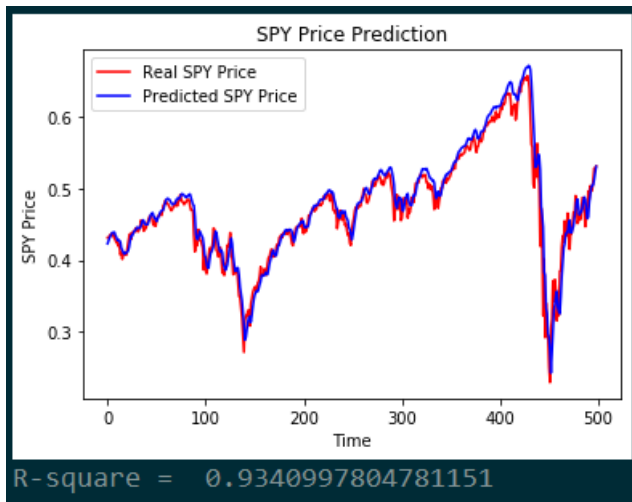


Figure: Output

效率市場假說(Efficient-market hypothesis, EMH)

- 強式效率 (Strong Form Efficiency):

目前股票價格充分反應了所有已公開和未公開之所有情報。雖然情報未公開，但投資者能利用各種管道來獲得資訊，所以，所謂未公開的消息，實際上是已公開的資訊且已反應於股票價格上。此種情形下，投資者也無法因擁有某些股票內幕消息而獲取高額報酬。

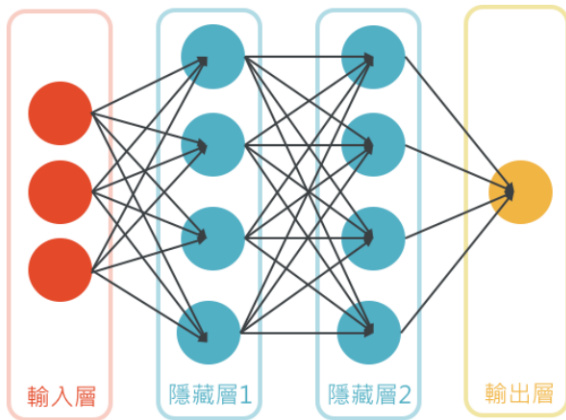


Figure: DNN structure