Retail and Marketing Analytics

Session 1

Gokhan Yildirim


WELCOME

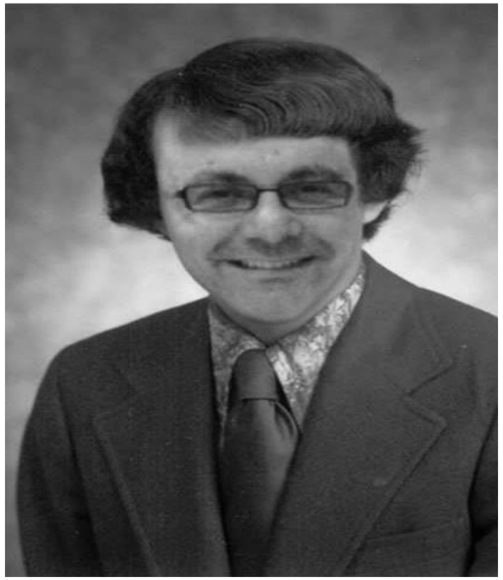
About me

Gokhan Yildirim

Associate Professor of Marketing

Expertise: Marketing Analytics

- Marketing
- Data
- Models



Office: 386C

E-mail: g.yildirim@imperial.ac.uk

Twitter: @gyildiri1

TAs

Britney Wang, PhD

Research interests:

- Substantive: digital marketing; marketing effectiveness
- Methodological: Time series modelling

E-mail: wanxin.wang13@imperial.ac.uk

Shameena Bonomally

PhD in particle physics

- Use of machine learning techniques for event categorisation in experimental particle physics.

E-mail: shameena.bonomally16@imperial.ac.uk

Outline

- Course logistics
- Introduction to retail & marketing analytics
- Demand forecasting: MLR and NNs
- Workshop with R

What will we learn in this module?

$$\sqrt{u} = -\frac{1}{2e^{-\rho t}}\left(\frac{-C_1 e^{-\sqrt{\gamma_{43}\gamma_{34}}t} + C_2 e^{\sqrt{\gamma_{43}\gamma_{34}}t}}{\sqrt{\gamma_{43}\gamma_{34}}} - e^{-\rho t}\frac{(p-c)}{\rho^2 - \gamma_{43}\gamma_{34}}\right)(\gamma_{41}\beta_1 + \gamma_{42}\beta_2 + \gamma_{43}\beta_3)$$

$$+\frac{1}{2e^{-\rho t}} C_4\left(\beta_1 + \beta_2\frac{\gamma_{42}}{\gamma_{41}} + \beta_3\frac{\gamma_{43}}{\gamma_{41}}\right)$$

$$\sqrt{u} = \frac{1}{2e^{-\rho t}}\left(\frac{C_4}{\gamma_{41}} + \frac{C_1 e^{-\sqrt{\gamma_{43}\gamma_{34}}t} \ldots}{\sqrt{\gamma_{43}\gamma_{34}}} \text{[illegible]}\, e^{-\rho t}\frac{(p-c)}{\rho^2 - \gamma_{43}\gamma_{34}}\right)(\gamma_{41}\beta_1 + \gamma_{42}\beta_2$$

$$+ \gamma_{43}\beta_3)$$

$$\sqrt{u} = \left(\frac{1}{2e^{-\rho t}\sqrt{\gamma_{43}\gamma_{34}}}\left(2C_2 e^{\sqrt{\gamma_{43}\gamma_{34}}T} + e^{-\rho T}\frac{(p-c)}{(\rho + \sqrt{\gamma_{43}\gamma_{34}})}\right)\right.$$

$$\left. - \frac{e^{-\rho T}\frac{\rho(p-c)}{\rho^2 - \gamma_{43}\gamma_{34}} + 2C_2 e^{\sqrt{\gamma_{43}\gamma_{34}}t}}{\sqrt{\gamma_{43}\gamma_{34}}} + e^{-\rho t}\frac{(p-c)}{\rho^2 - \gamma_{43}\gamma_{34}}\right)(\gamma_{41}\beta_1 + \gamma_{42}\beta_2$$

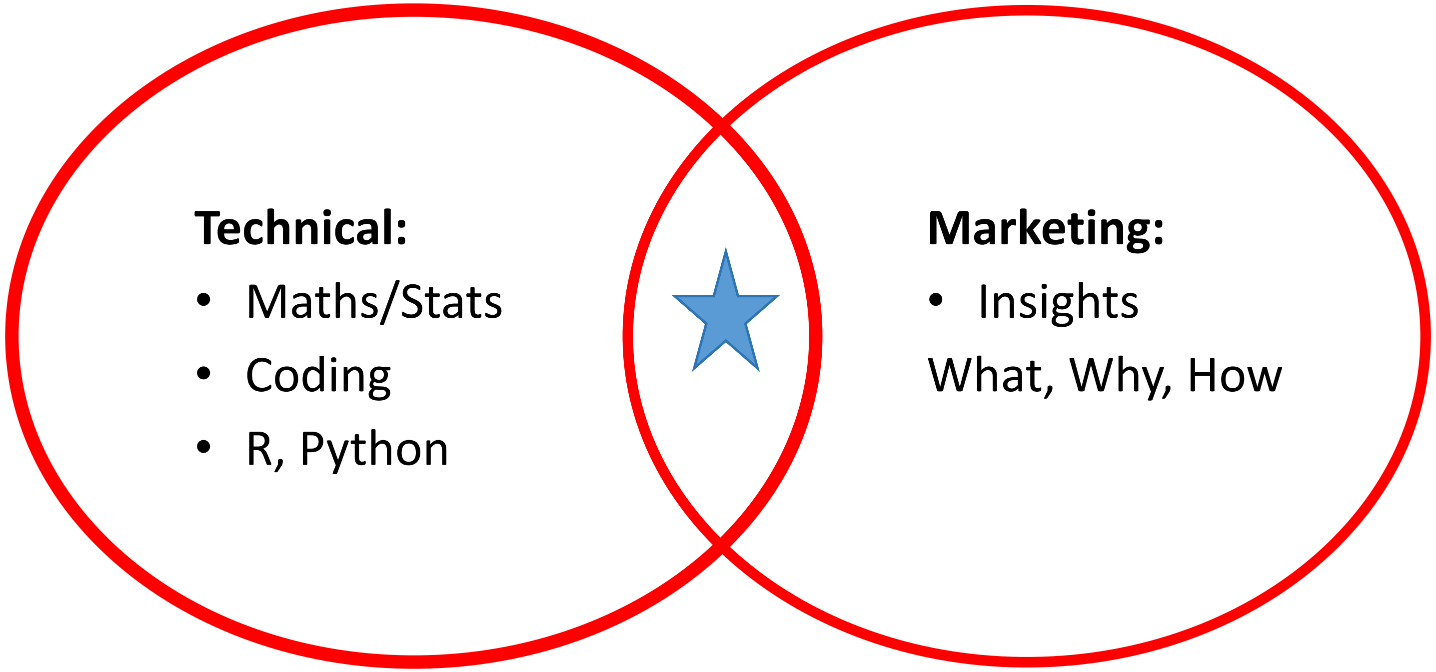
$$+ \gamma_{43}\beta_3)$$

What will we learn in this module?

Solving a range of retail and marketing problems with the help of applied quantitative models

Learning only techniques is not enough…

YOUR SKILLSETS



Your learning experience

1. Lectures on key marketing models/methods

2. Software applications using real-life data

3. Simulation game

Module Schedule

SESSION 1: Demand Forecasting	SESSION 2: Retail Promotions and Advertising	SESSION 3: Marketing Resource Allocation	SESSION 4: Winning Hearts, Minds and Sales	SESSION 5: The Future of Retailing
Lecture: • Introduction to RMA • Multiple Linear Regression and Neural Networks **R tutorial:** Demand forecasting with Neural Networks	**Lecture:** • Retail Promotions and SCAN*PRO Model • Advertising Effectiveness and Adstock Model **R tutorial:** SCAN*PRO and AdStock Models **Game:** Data Analytics Simulation: Strategic Decision Making (1st round)	**Lecture:** • Modelling omni-channel marketing **R tutorial:** Resource allocation with VAR models **Game:** Data Analytics Simulation: Strategic Decision Making (2nd round)	**Lecture:** • Customer Attitudinal Metrics **R Tutorial:** Mind-set Metrics for Guiding Marketing Mix Decisions **Game:** Data Analytics Simulation: Strategic Decision Making (3rd and 4th round)	**Lecture:** Digital marketing Strategies How AI is re-shaping retailing **Lecture:** Big Data and Decision Making Game debriefing Wrap-up

Assessment

Class Participation
10%

Individual Coursework
50%

Report
(50%)

Group Coursework
(Simulation Game)
50%

Game performance
(20%)

Report
(20%)

Individual coursework

Apply at least one of the methods/models covered in the course to a dataset of their own choice in order to solve a particular marketing problem or demonstrate a marketing opportunity (e.g. demand forecasting, promotion/advertising effectiveness, budget allocation).

- Background to the marketing problem/opportunity
- Application of relevant models
- What implications are drawn from the analyses?
- Quality of overall structure

Some online data resources

- Kaggle (https://www.kaggle.com/datasets).
  - https://www.kaggle.com/c/rossmann-store-sales
  - https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting
  - https://www.kaggle.com/c/competitive-data-science-predict-future-sales/data

- Dunnhumby

https://www.dunnhumby.com/careers/engineering/sourcefiles

- UCI Machine Learning Repository:

https://archive.ics.uci.edu/ml/datasets.php?format=&task=&att=&area=bus&numAtt=&numIns=&type=&sort=nameUp&view=table

- INFORMS- Tablet Computer Dataset:

https://business.uc.edu/tablet-computer-data.html

Group coursework

Simulation Game

- You will act as a **brand manager** for a laundry detergent brand
- Your task is to improve the brand's performance by **leveraging data** to determine the best marketing strategy
- You will have to **make strategic decisions** about product composition, predict demand, set prices, and determine promotional spending, while communicating your strategy effectively to your managers
- The game makes use of **real-life consumer data** from a multinational consumer goods company

Group coursework

Game performance:

– Cumulative profit

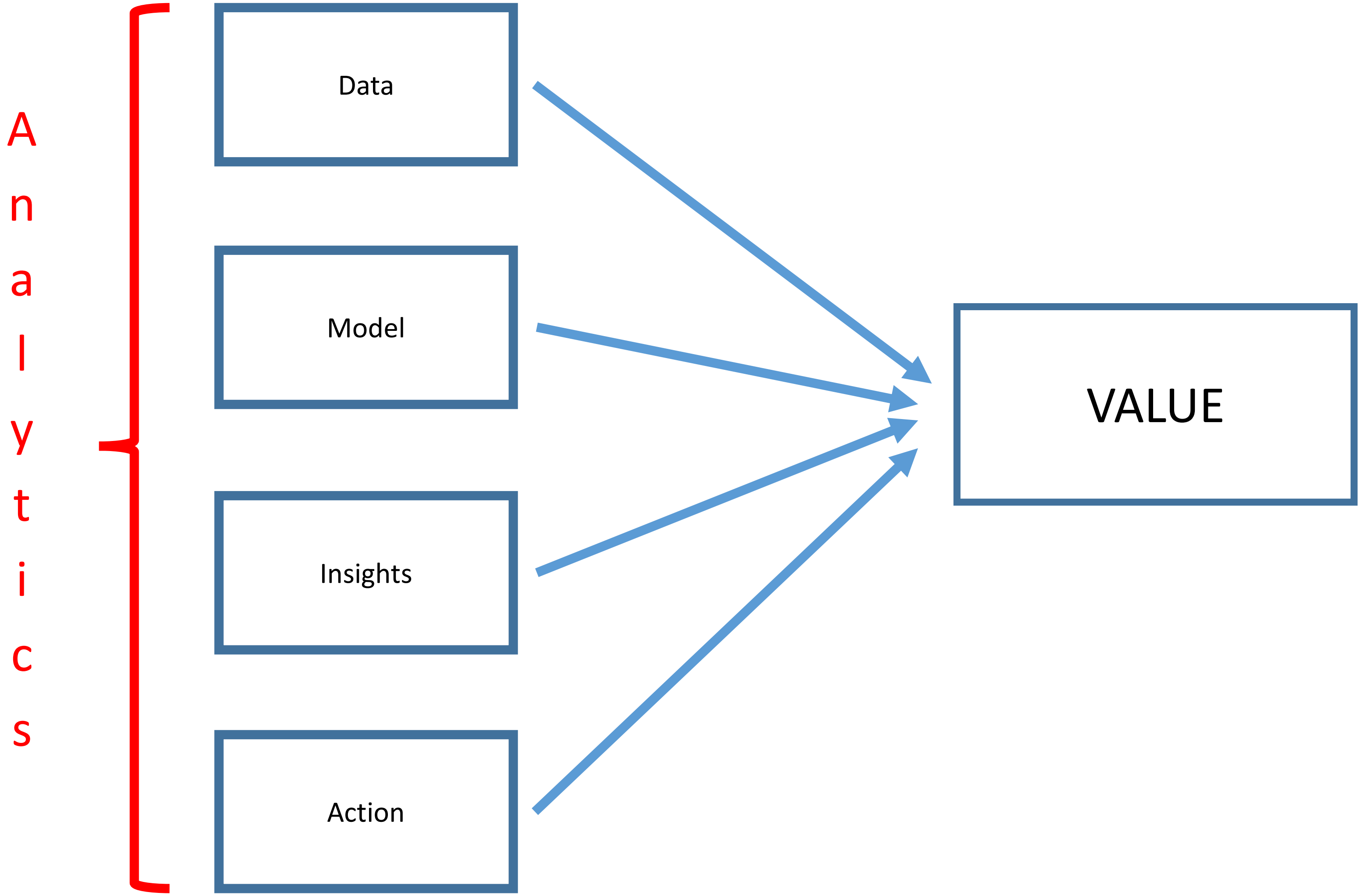
– Cumulative revenue

– Final market share

E.g. If you rank 1st on cumulative profit, 3rd on cumulative revenue, and 5th on market share, your overall performance score will be 3 ((1+3+5)/3=3).

Reflection report:

What is marketing analytics about?



Why Learn About Marketing Analytics

5%

Of managers in retail use analytics to make pricing decisions.

Marketing Analytics (Grigsby 2018)

Why Learn About Marketing Analytics

$1 trillion

“Marketers are always asking for more money, but can rarely explain how much incremental business this money will generate”

↕

Accountability drives the marketing department’s influence within the firm

The dawn of marketing’s next golden age: $200 billion and counting (McKinsey & Company);
Verhoef, P. C., & Leeflang, P. S. (2009). Understanding the marketing department's influence within the firm, Journal of Marketing, 73(2), 14-37

Demand Forecasting

Learning objectives

Define “**demand forecasting**” and explain why it is important in retail

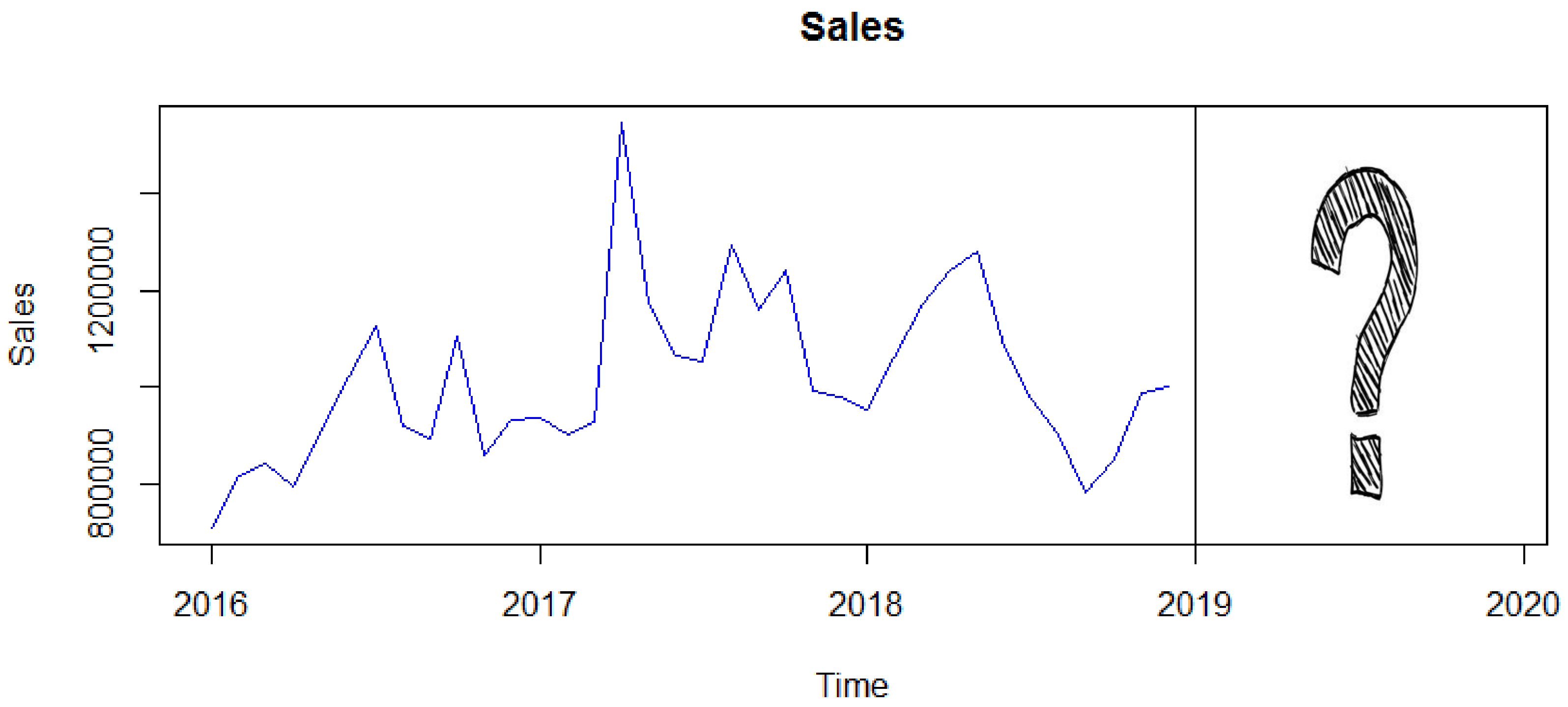
Discuss **alternative forecasting approaches** for retail sales forecasting

Understand **neural networks** and apply the model to forecast retail sales

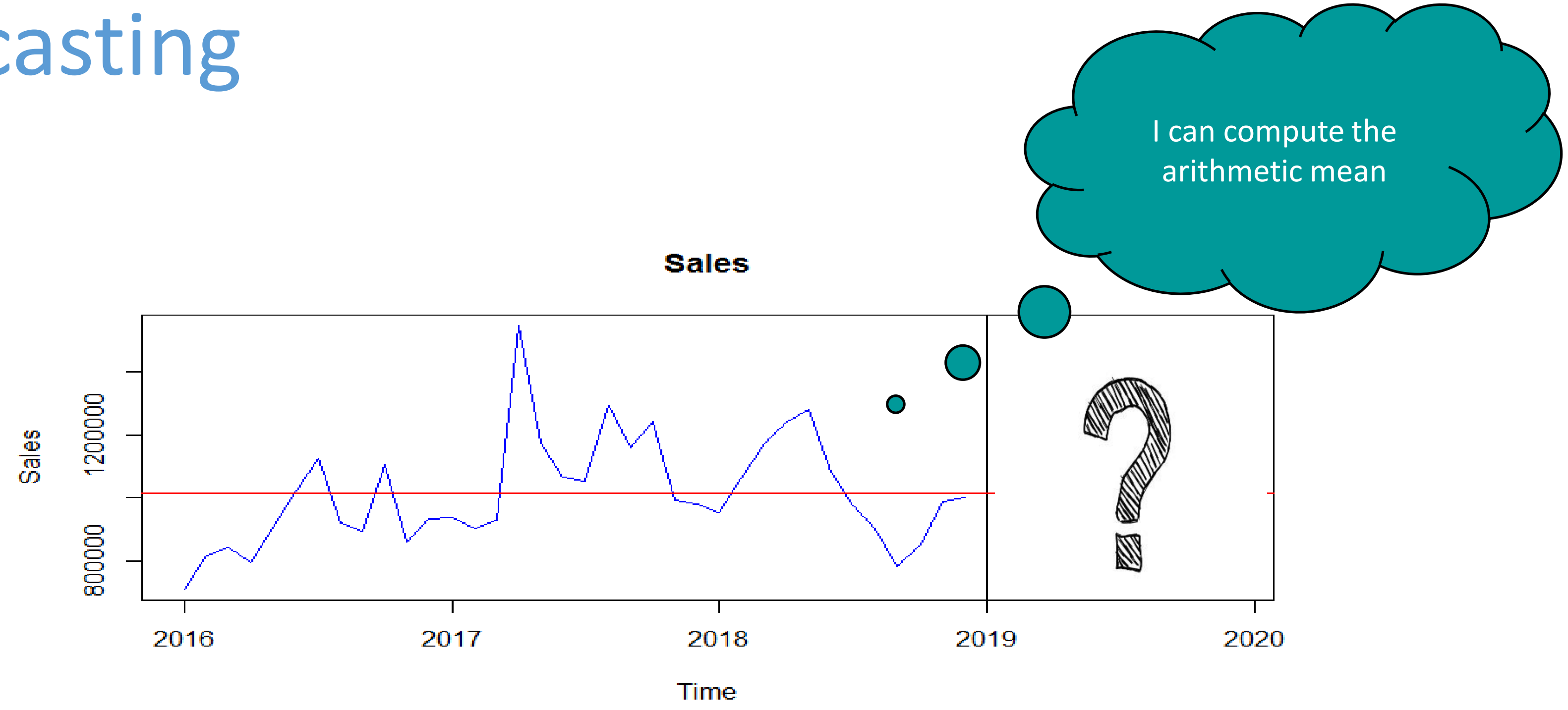
What is Forecasting?

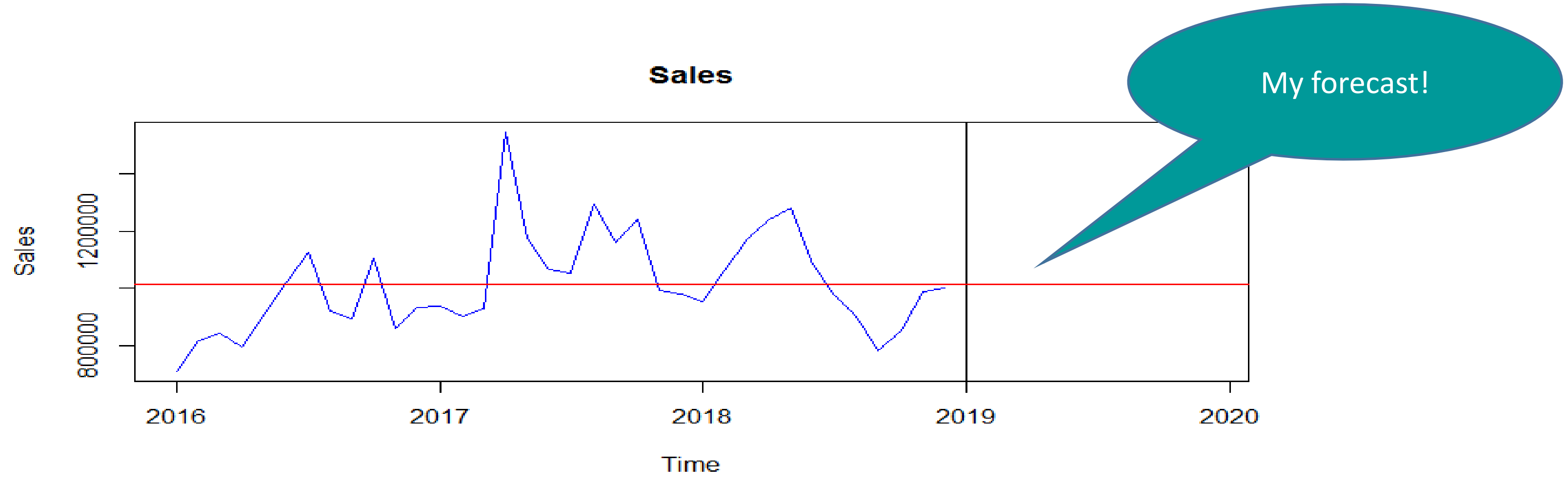
WHAT

What will happen next?



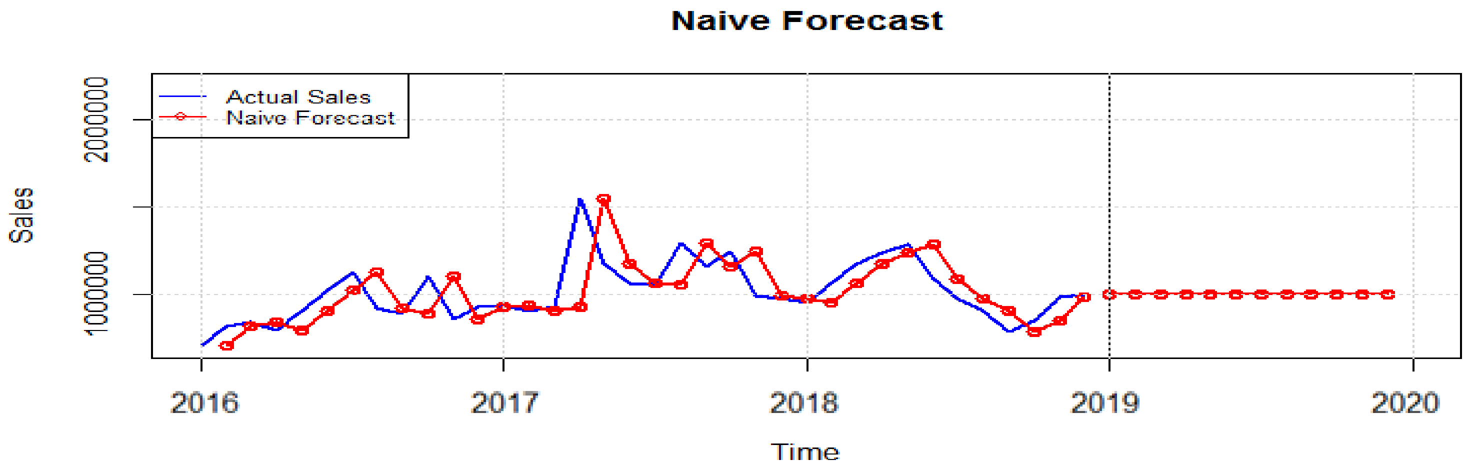
Sales forecasting



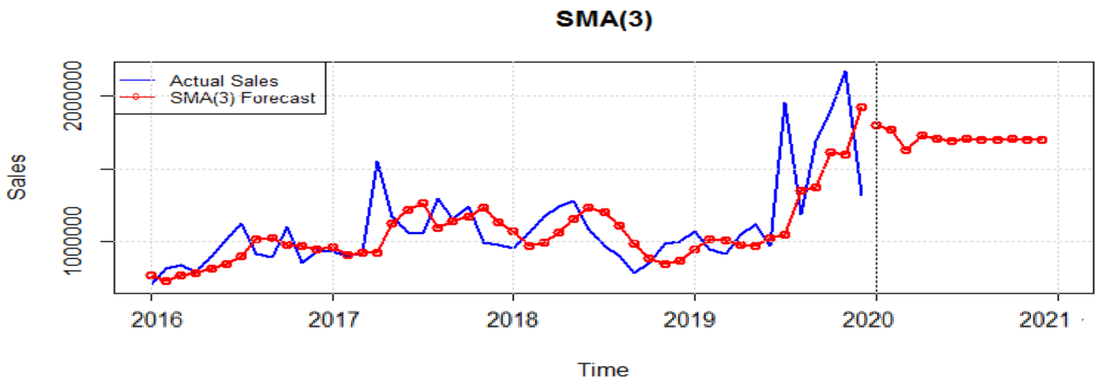


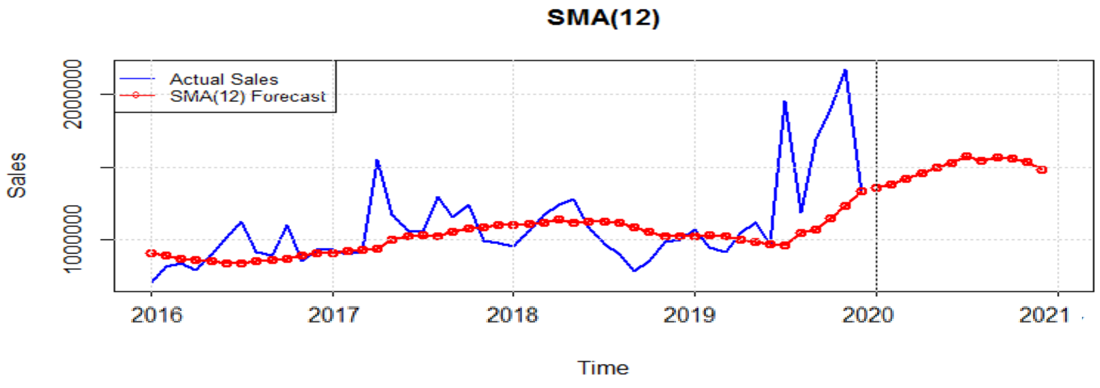
Alternative forecast methods?

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



Alternative forecast methods





Simple Moving Average?

$$\hat{y}_{t+1|t} = \frac{1}{k} \sum_{i=t-k+1}^{t} y_i$$

- SMA calculates the average of last *k* observations as a forecast for the next period.

- *m* refers to length of SMA.

Alternative forecast methods?

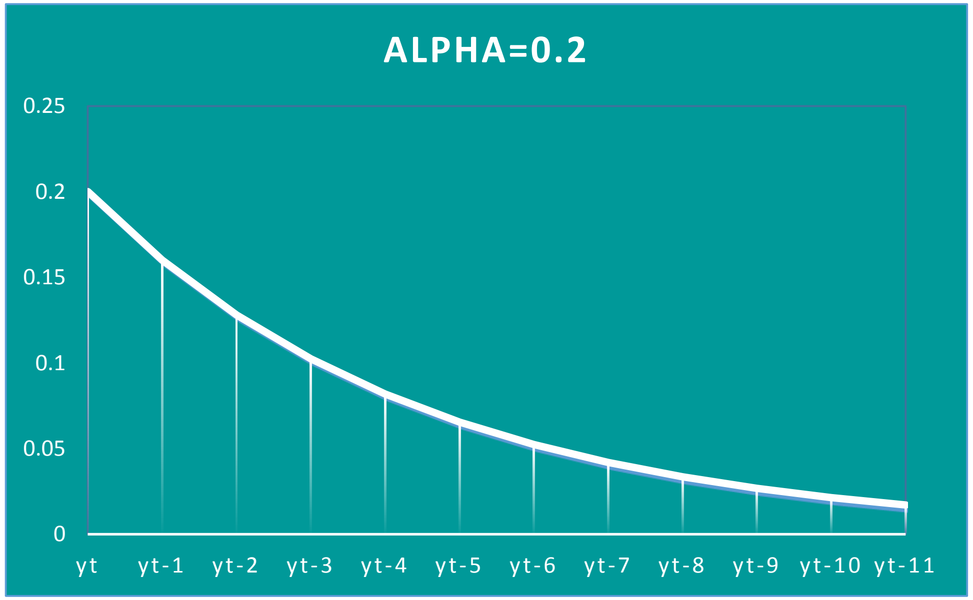
Simple Exponential Smoothing

$$\hat{y}_{t+1|t} = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \ \ldots$$

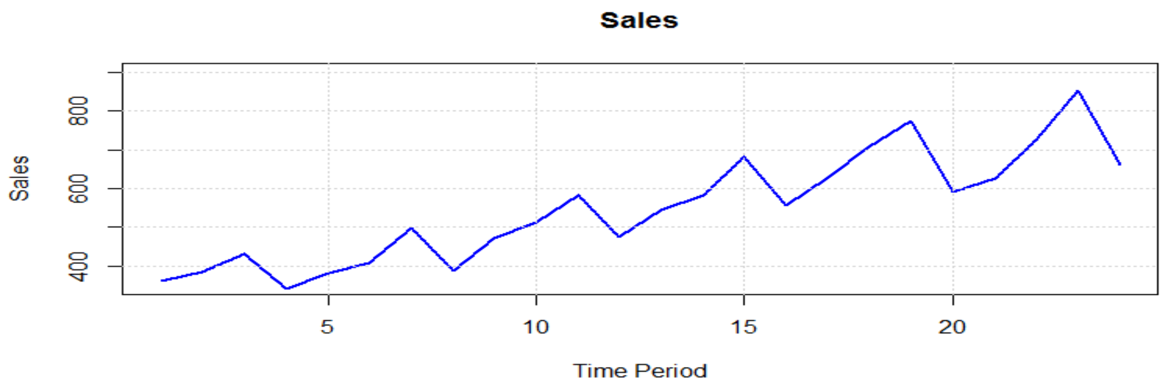
$$\Downarrow$$

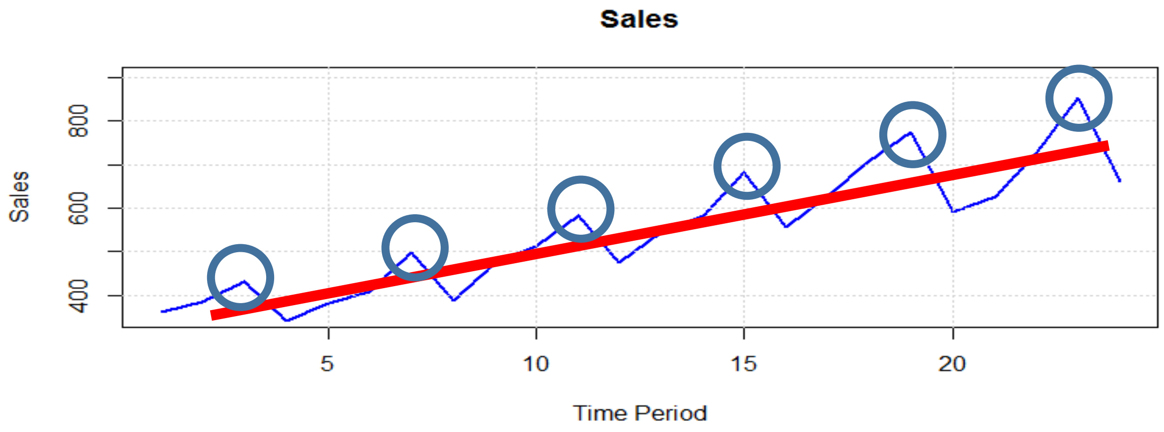
$$\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1}$$

Observation	α **= 0.2**
y_t	0.2
y_{t-1}	$(0.2) \times (0.8)$
y_{t-2}	$(0.2) \times (0.8)^2$
y_{t-3}	$(0.2) \times (0.8)^3$
y_{t-4}	$(0.2) \times (0.8)^4$
y_{t-5}	$(0.2) \times (0.8)^5$



What if there are seasonality and trend patterns?



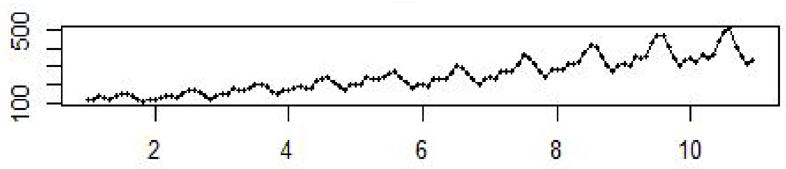


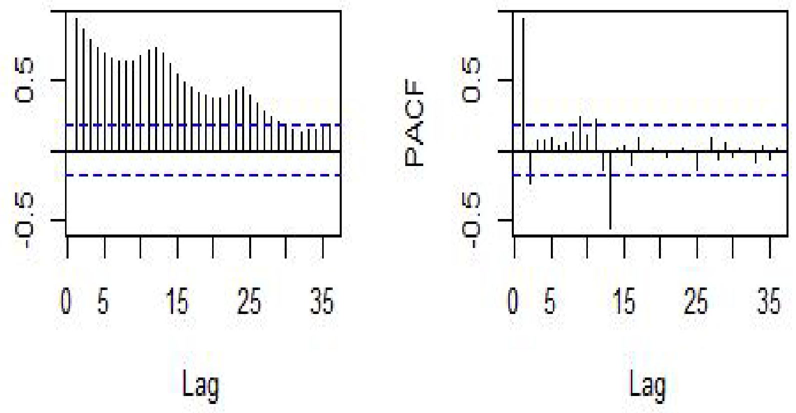
ARIMA Models

ARIMA Model:

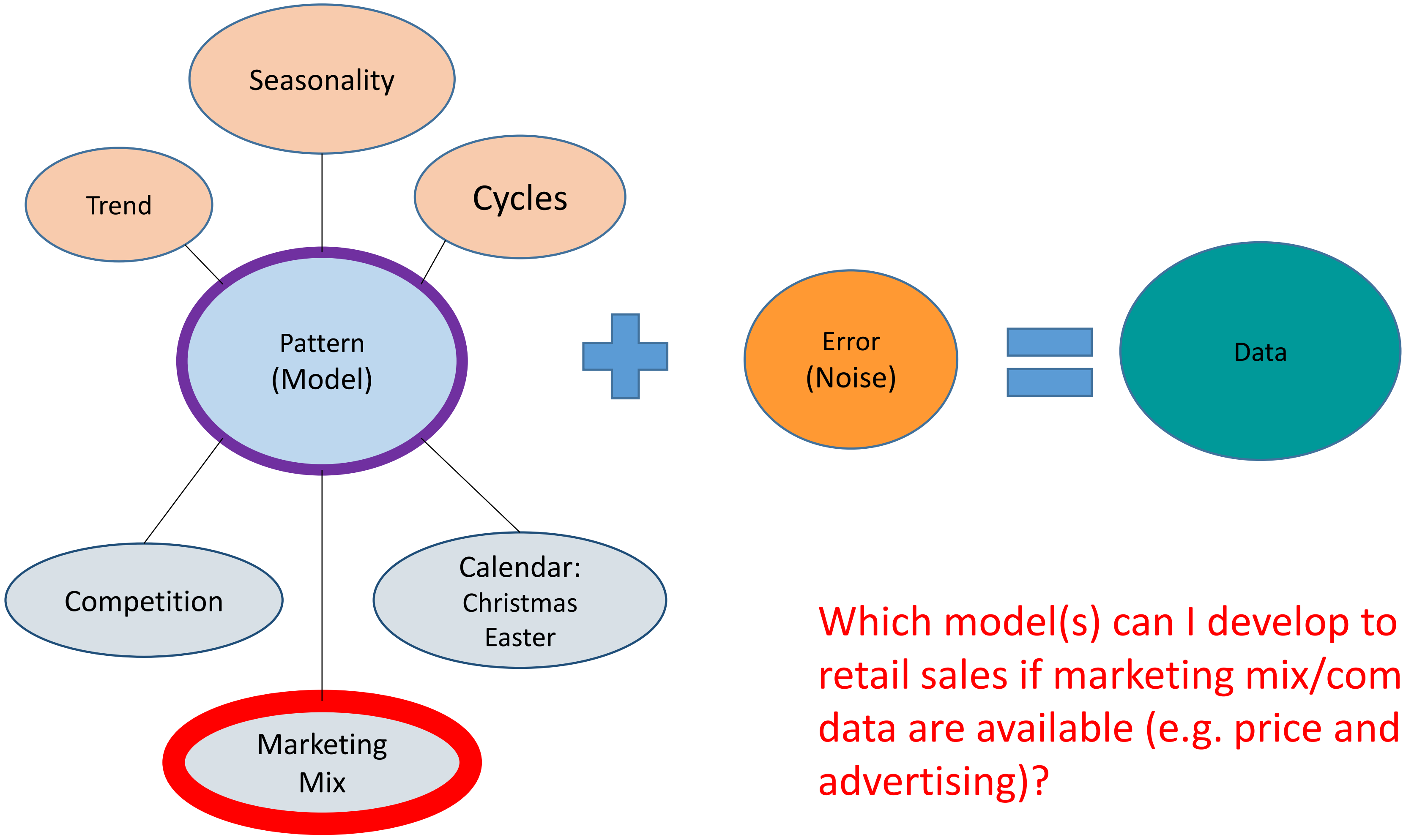
$$\left(1 - \sum_{p=1}^{p} \phi_i \, L^i \right) Y_t = \left(1 + \sum_{q=1}^{q} \theta_i \, L^i \right) \varepsilon_t$$

- Non-stationarity
- Past dynamics of Sales (AR component)
- Past dynamics of forecast errors (MA component)
- Seasonality → SARIMA model





Data patterns



Which model(s) can I develop to forecast retail sales if marketing mix/competition data are available (e.g. price and advertising)?

One variable predicted by others → $y = g(x_1, \dots, x_k) + Noise$

$\Downarrow$

Forecasting with Multiple Regression

$$Sales_t = c + \beta Adv_t + \gamma Price_t + \delta Promo_t + \epsilon_t, \qquad \epsilon \sim i.i.d.\ \mathrm{N}(0, \sigma_\epsilon^2)$$

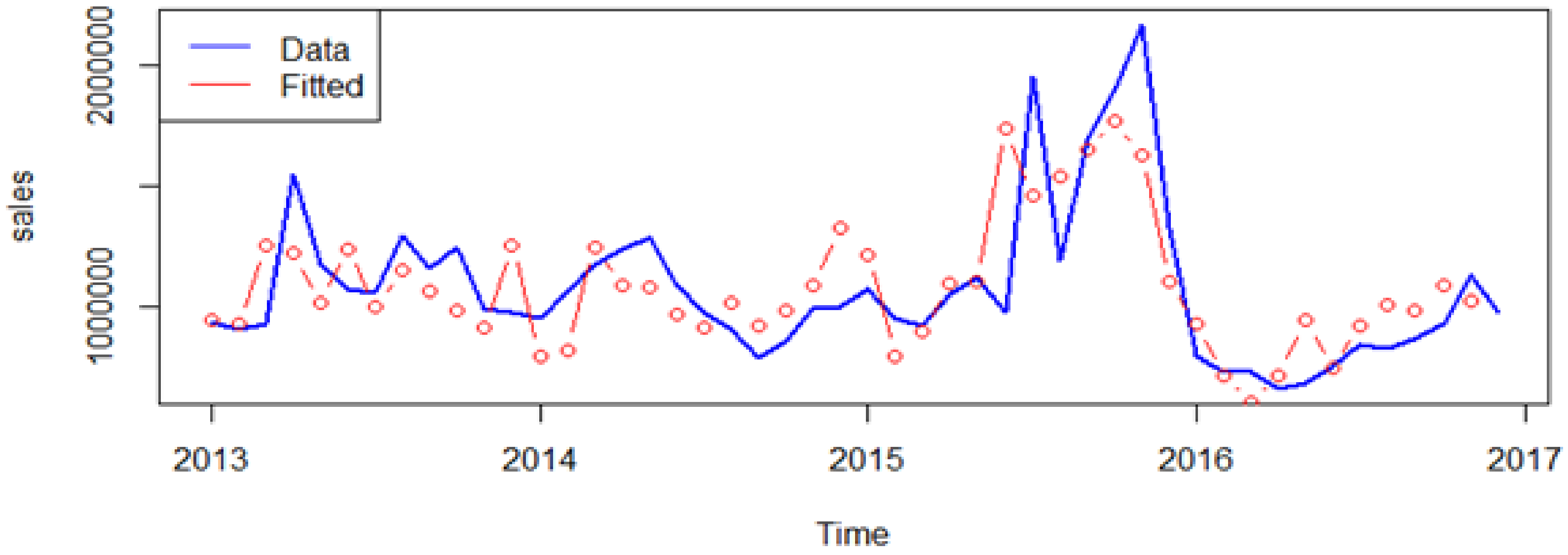
$$Sales_t = c + \beta Adv_t + \gamma Price_t + \delta Promo_t + \boxed{\Theta D_t} + \epsilon_t$$

$$Sales_t = c + \boxed{\alpha Sales_{t-1}} + \beta Adv_t + \gamma Price_t + \delta Promo_t + \Theta D_t + \epsilon_t$$

$$Sales_t = c + \alpha Sales_{t-1} + \beta Adv_t + \gamma Price_t + \delta Promo_t + \Theta D_t$$
$$+ \boxed{\psi Adv_t Promo_t} + \epsilon_t$$

Multiple Linear Regression

$$\widehat{Sales}_t = 5.4e + 5 + 0.3\, Sales_{t-1} + 862\, Adv_t - 100.2\, Price_t + 40.3\, Promo_t$$
$$+ 31.4\, Adv_t Promo_t + \widehat{\Theta} D_t$$

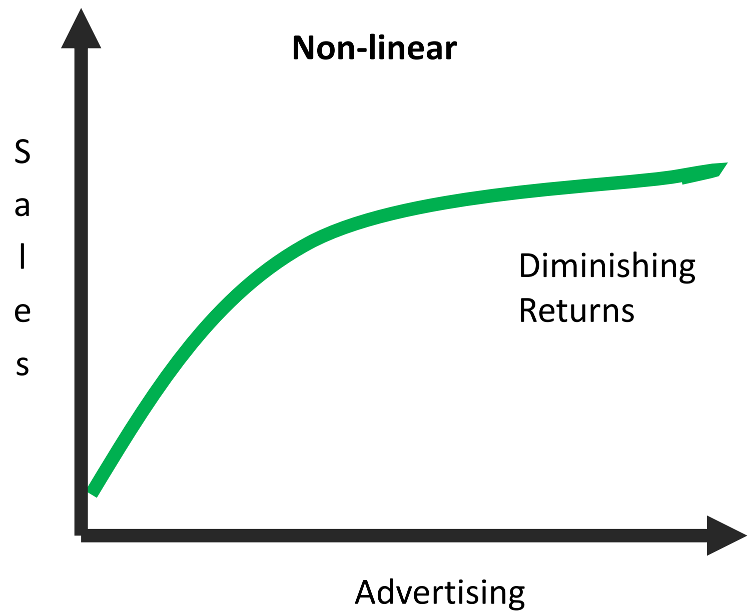


Activity

1. Can you plot the ad spend vs. the total sales?
2. Can you calculate the average sales?
3. Can you calculate the marginal sales?
4. **What do you conclude?**

Ad spend	Total Sales	Average Sales	Marginal Sales
0			
100	900		
200	1600		
300	2300		
400	2500		
500	2500		
600	2400		

Diminishing returns

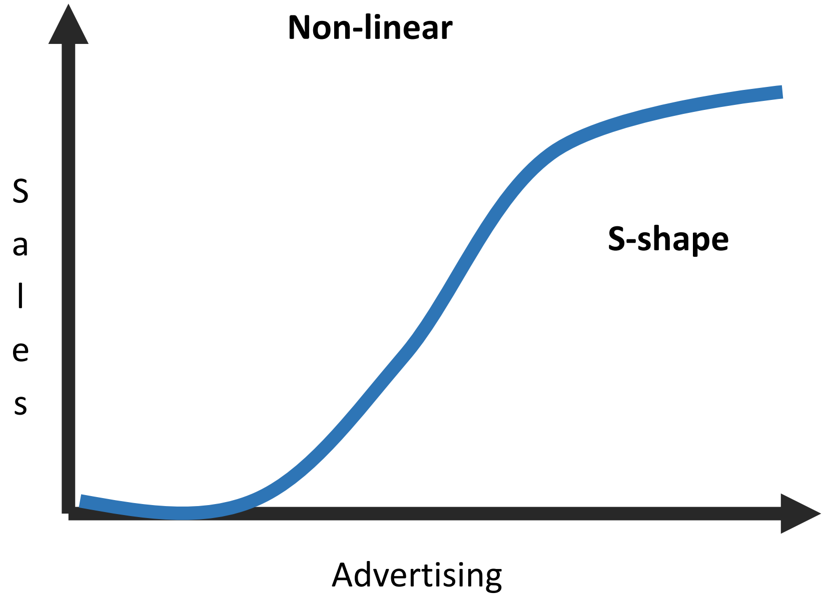


How can we introduce this type of non-linearity in the model?

$$y = c + \beta \ln(Adv)\,, \qquad \beta > 0$$

$$\ln(y) = c + \beta \ln(Adv), \qquad 0 < \beta < 1$$

S-shape

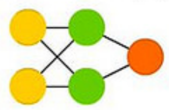


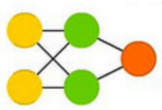
How can we introduce this type of non-linearity in the model?

$$Sales = a\left(\frac{1}{1 + e^{-(b*Adv+\mu_{adv})}}\right)$$

a curve's maximum value
b shows the steepness of the curve

Neural Networks





Neural Networks

0

1

2

$\boxed{3}$

4

5

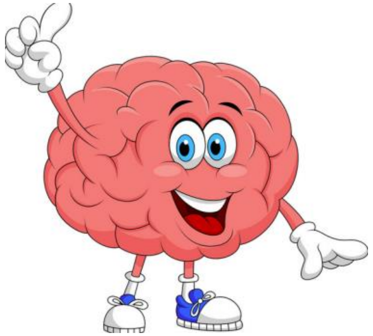
6

7

8

9

What if I told you to write a program that tells what it thinks the digit is?



Neural Network

What are the neurons?

How are they connected?

25

25

0.01

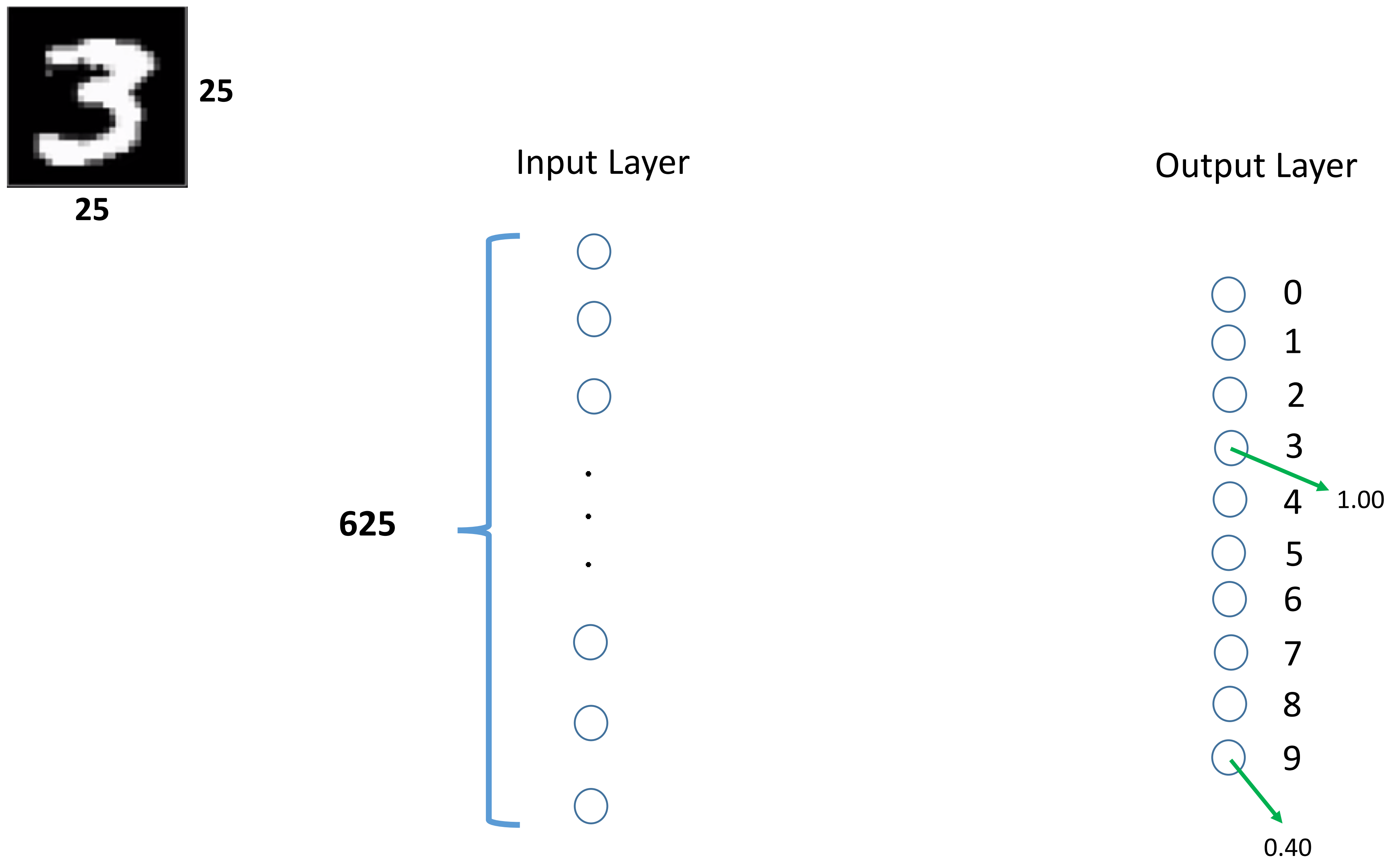
Activation

0.56

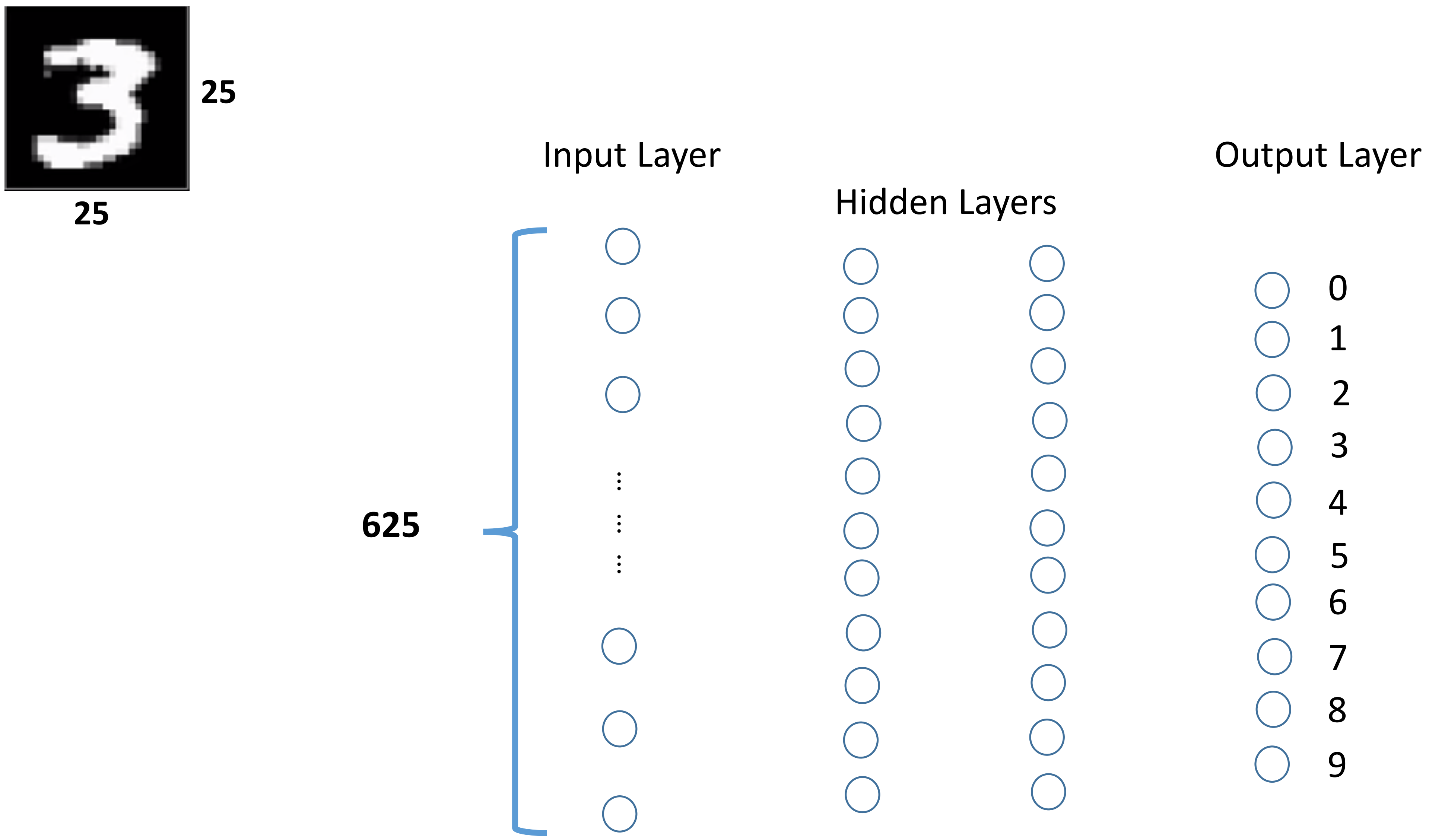
Activation

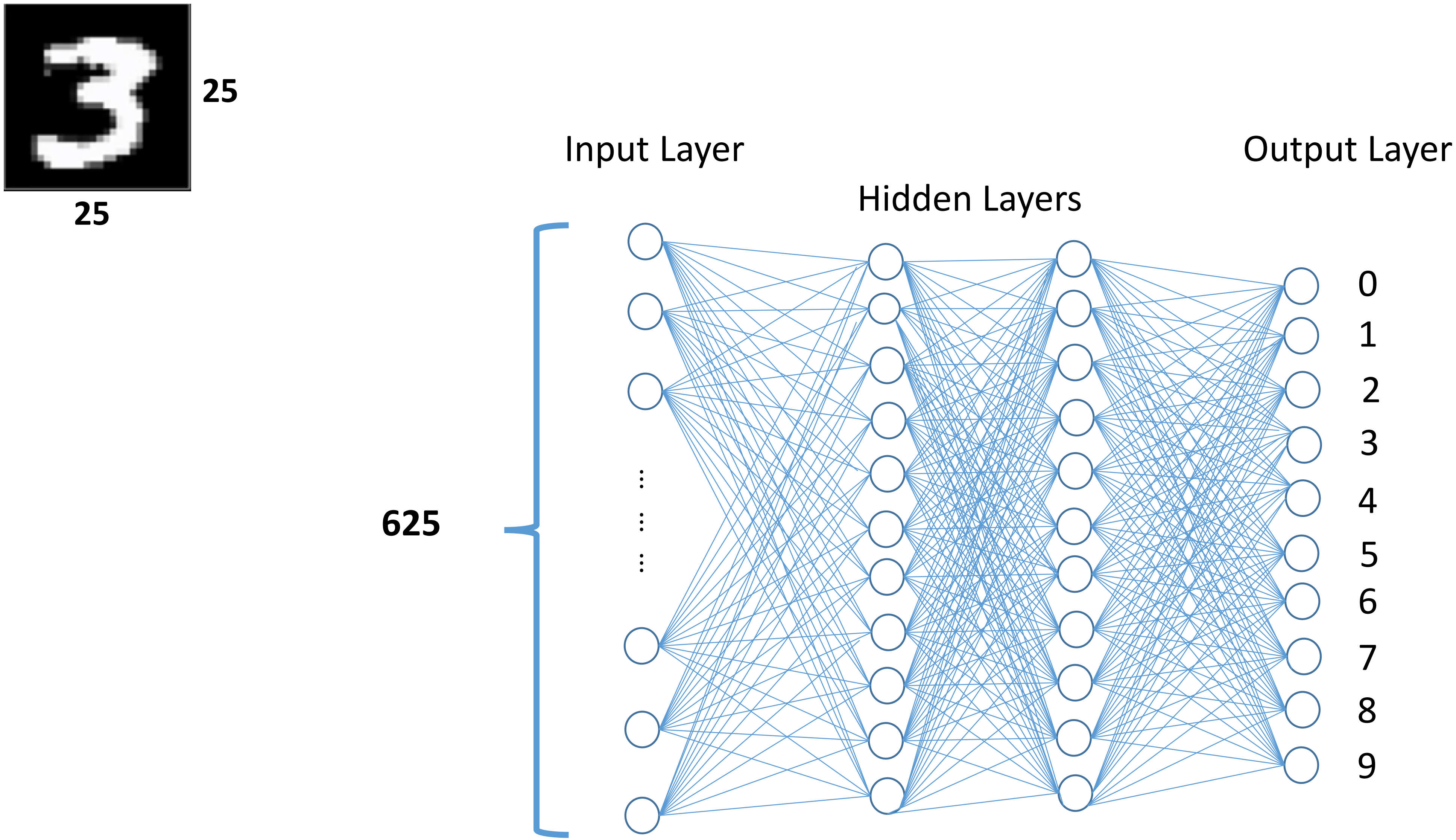
1.00

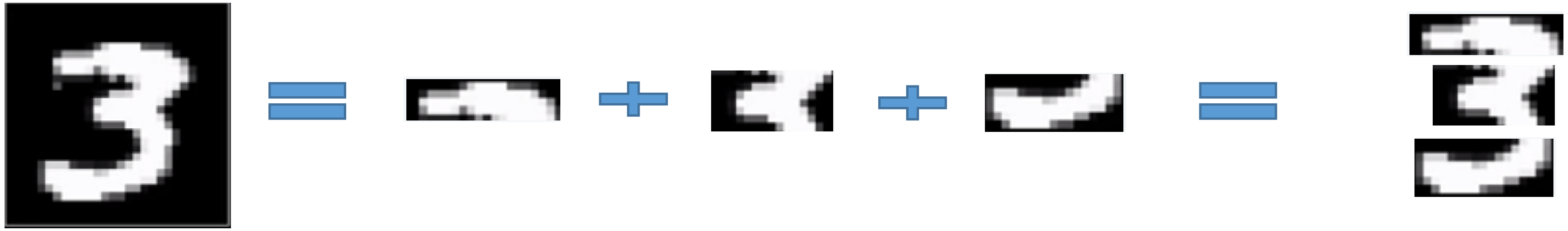
Activation

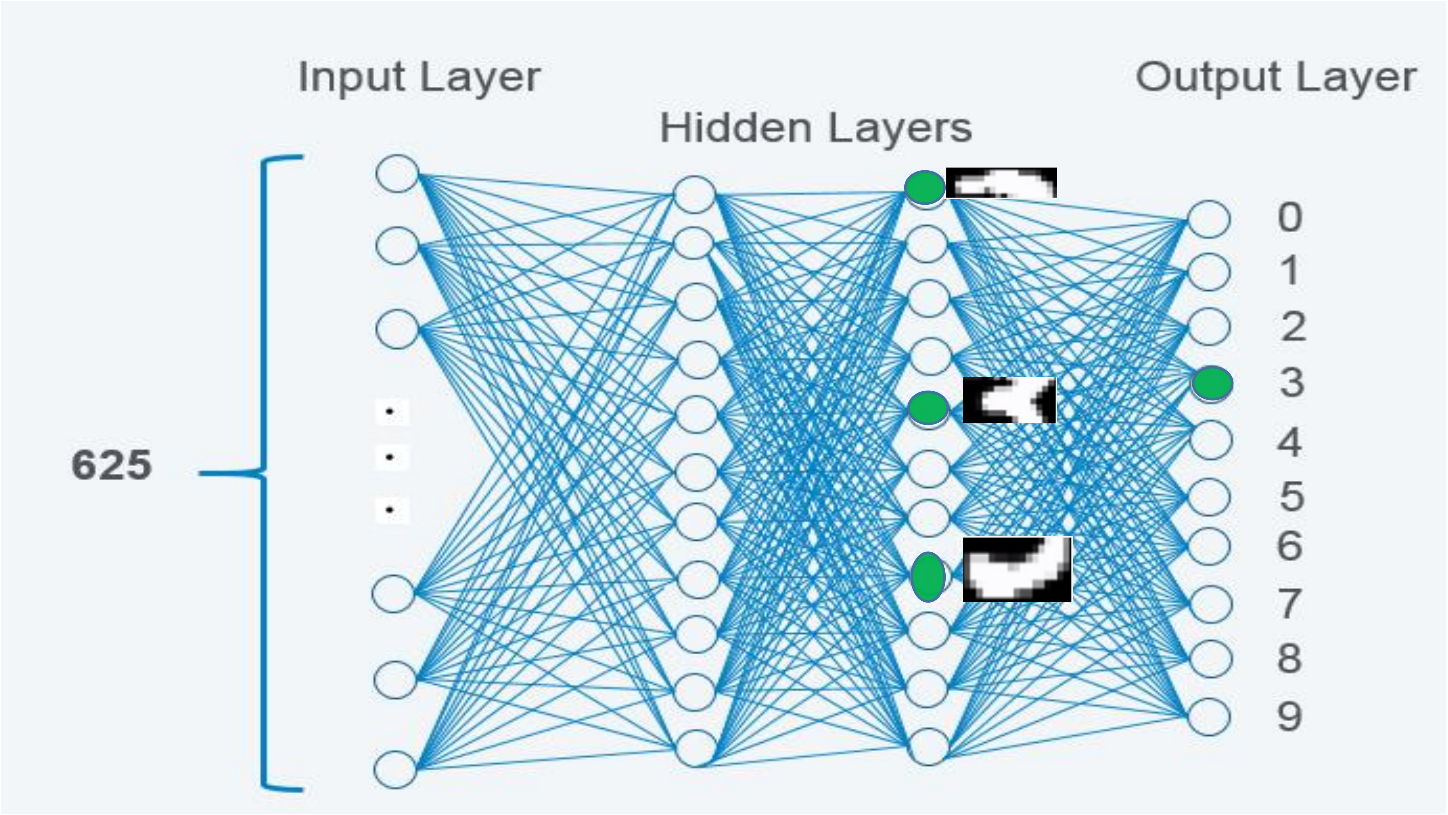


25
25
Input Layer
Output Layer
625
0
1
2
3
4
5
6
7
8
9
1.00
0.40


25
25
Input Layer
Hidden Layers
Output Layer
625
0
1
2
3
4
5
6
7
8
9


25
25
Input Layer
Hidden Layers
Output Layer
625
0
1
2
3
4
5
6
7
8
9

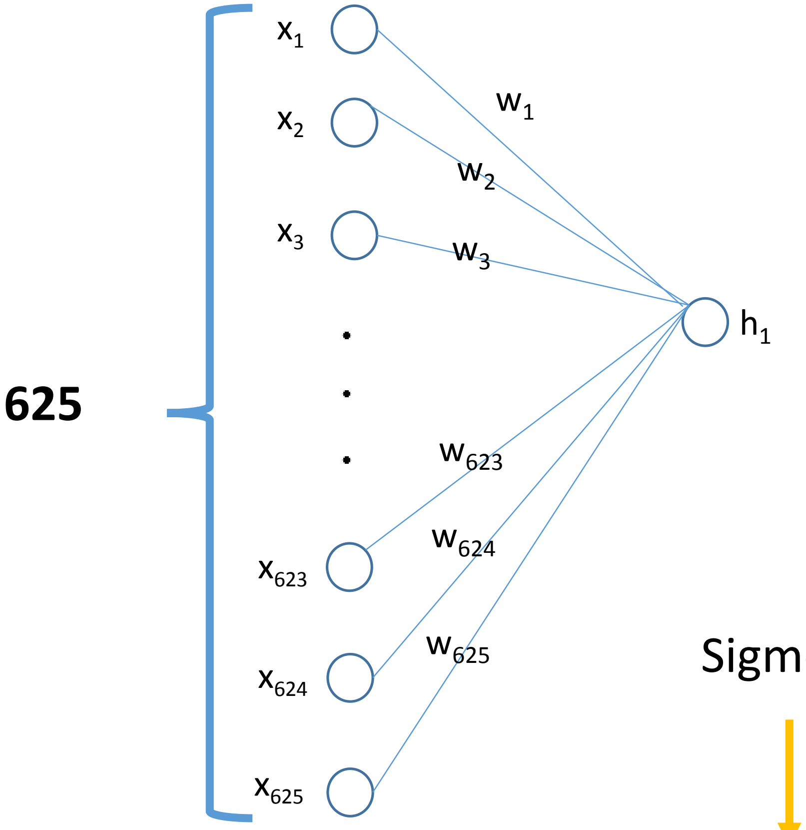




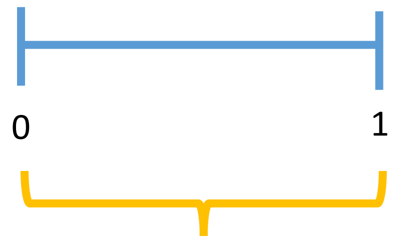
Input Layer
Hidden Layers
Output Layer
625
0
1
2
3
4
5
6
7
8
9

- WEIGHTS
- BIASES
- ACTIVATION FUNCTION

Input Layer



$$h_1 = \overbrace{w_1x_1 + w_2x_2 + \ldots + w_nx_n}^{\textbf{Weighted Sum}} + \underbrace{b_1}_{\textbf{Bias}}$$



Sigmoid

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

$$h_1 = \sigma(w_1x_1 + w_2x_2 + \ldots + w_nx_n + b_1)$$

From one layer to the next...

In matrix form:

$$\sigma\left(\underbrace{\begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{k1} & \cdots & w_{kn} \end{bmatrix}}_{\text{Weights matrix}} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}}_{\text{Input vector}} + \underbrace{\begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix}}_{\text{bias vector}}\right)$$

In a compact form:

$$\sigma(\mathbf{W}\mathbf{x} + \mathbf{b})$$

How many parameters?

weights

$$\underbrace{625 \times 11}_{\text{From input layer to hidden layer 1}} + \underbrace{11 \times 11}_{\text{From hidden layer 1 to hidden layer 2}} + \underbrace{11 \times 10}_{\text{From hidden layer 2 to output layer}}$$

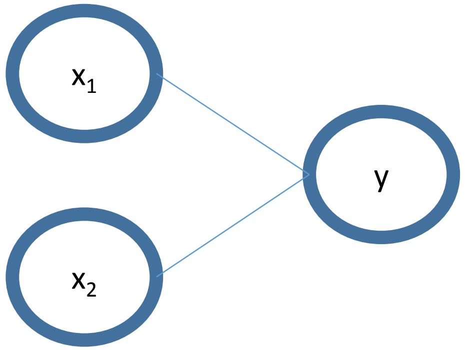
biases

$$\underbrace{11}_{\text{Hidden layer 1}} + \underbrace{11}_{\text{Hidden layer 2}} + \underbrace{10}_{\text{Output layer}}$$

IN TOTAL 7138 parameters !

Example: clicked or not?

Click (Yes=1, No=0)	0	1	1	1	0	0	1	0	?
Duration	1.5	2	4.5	3	2	2	5	2	2.5
Pages	3	1	2	1	4	3	1	3	1



$$NN(x_1, x_2) = sigmoid(w_1 x_1 + w_2 x_2 + b)$$

$$NN(3,1.5) = sigmoid(w_1 3 + w_2 1.5 + b)$$

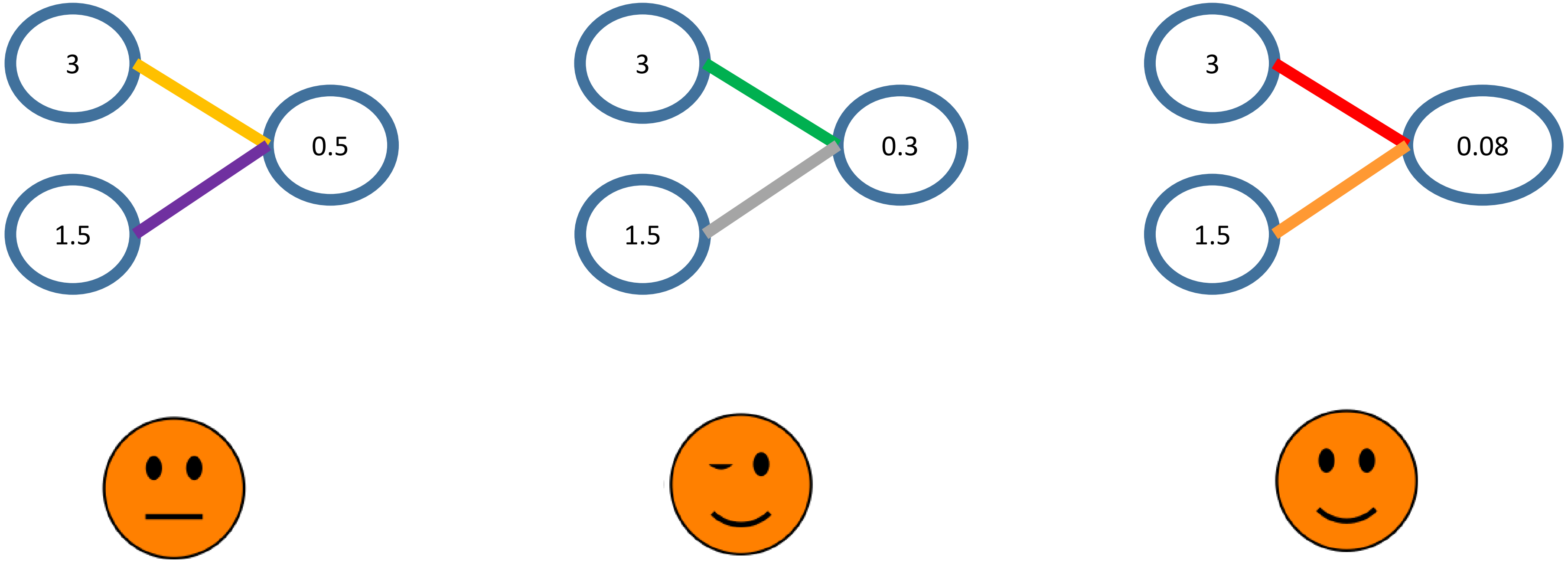
$$NN(3,1.5) = sigmoid(0.2 \times 3 + 0.5 \times 1.5 + 0.3)$$

$$NN(3,1.5) = sigmoid(1.65)$$

$$NN(3,1.5) = 0.84$$

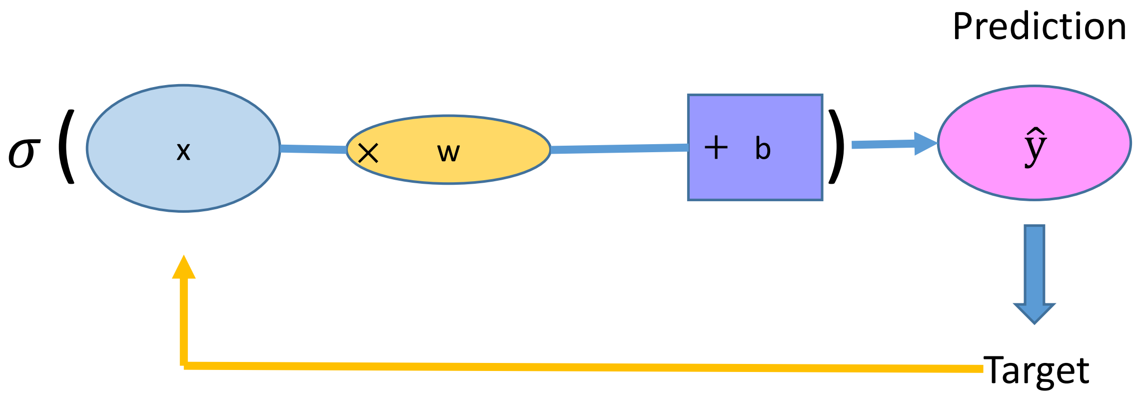
Change weights and bias parameters…

Click (Yes=1, No=0)	0	1	1	1	0	0	1	0	?
Duration	1.5	2	4.5	3	2	2	5	2	2.5
Pages	3	1	2	1	4	3	1	3	1



Two steps in training a neural network…

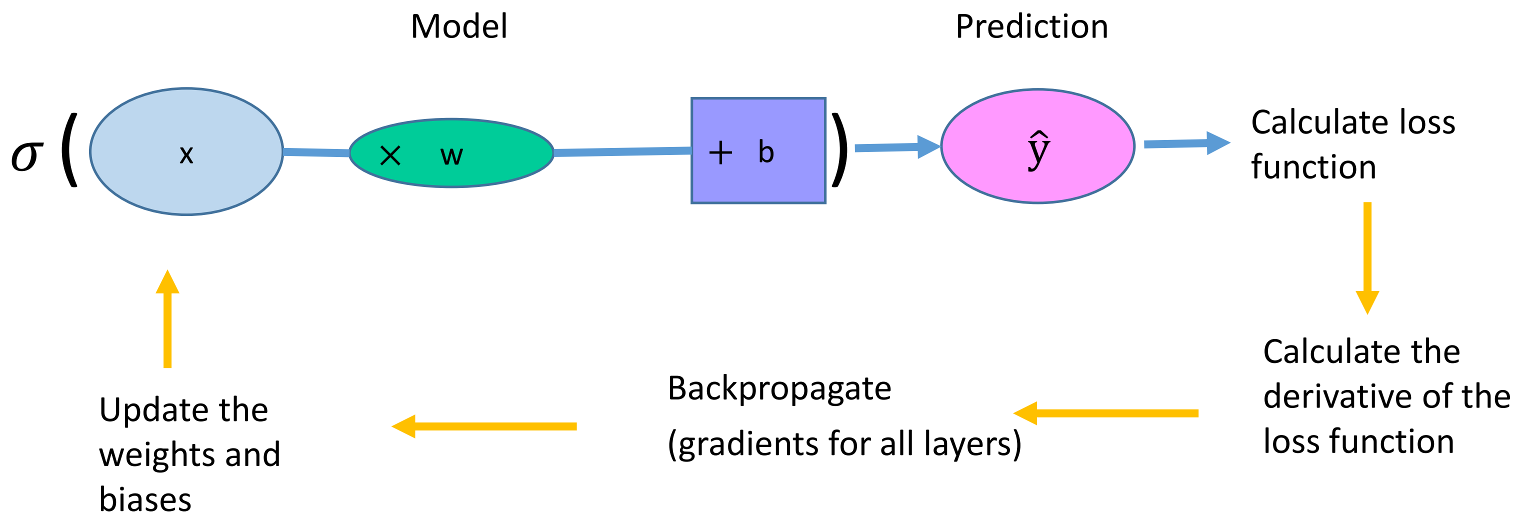
- Feedforward
- Backpropagation



Is your prediction close to the target?

Feedforward and Backpropagation

- Feedforward
- Backpropagation



Loss Function?

$$Error = \underbrace{Target}_{y} - \underbrace{Prediction}_{\hat{y}}$$

$$Squared\ Error = (y - \hat{y})^2$$

$$Sum\ of\ Squared\ Errors:$$

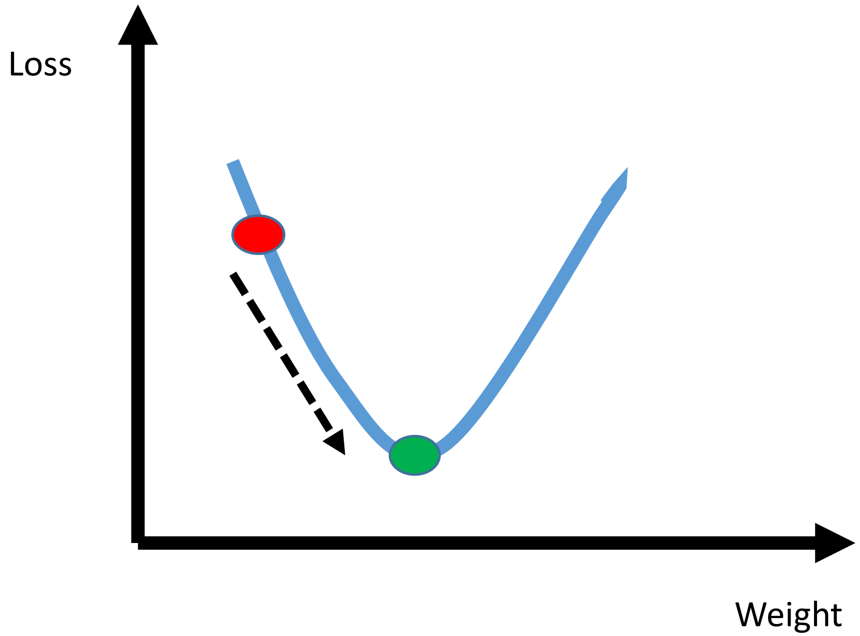
$$Loss = \sum_{i=1}^{n} (y_i - \widehat{y_i})^2$$

$$Objective?$$

$\min \sum_{i=1}^{n} (y_i - \widehat{y_i})^2$

Gradient Descent:

Derivative of the Loss Function with respect to weights and biases.



Neural Networks: summary

NN can be applied for two main categories:

- **Classification:** y is categorical (e.g. brand choice)
- **Regression:** y is continuous (e.g. sales volume)

Neural Networks: summary

Input layer, hidden layer(s), and output layer.

Neural network is all about weights, biases and activation function.

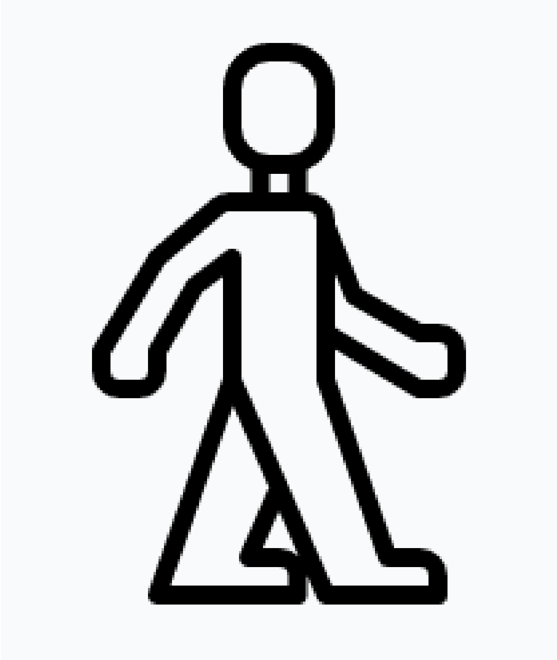
Computationally expensive.

Model set-up:

- How many hidden layers?
- How many nodes in each hidden layer?
- Iterations and Error level

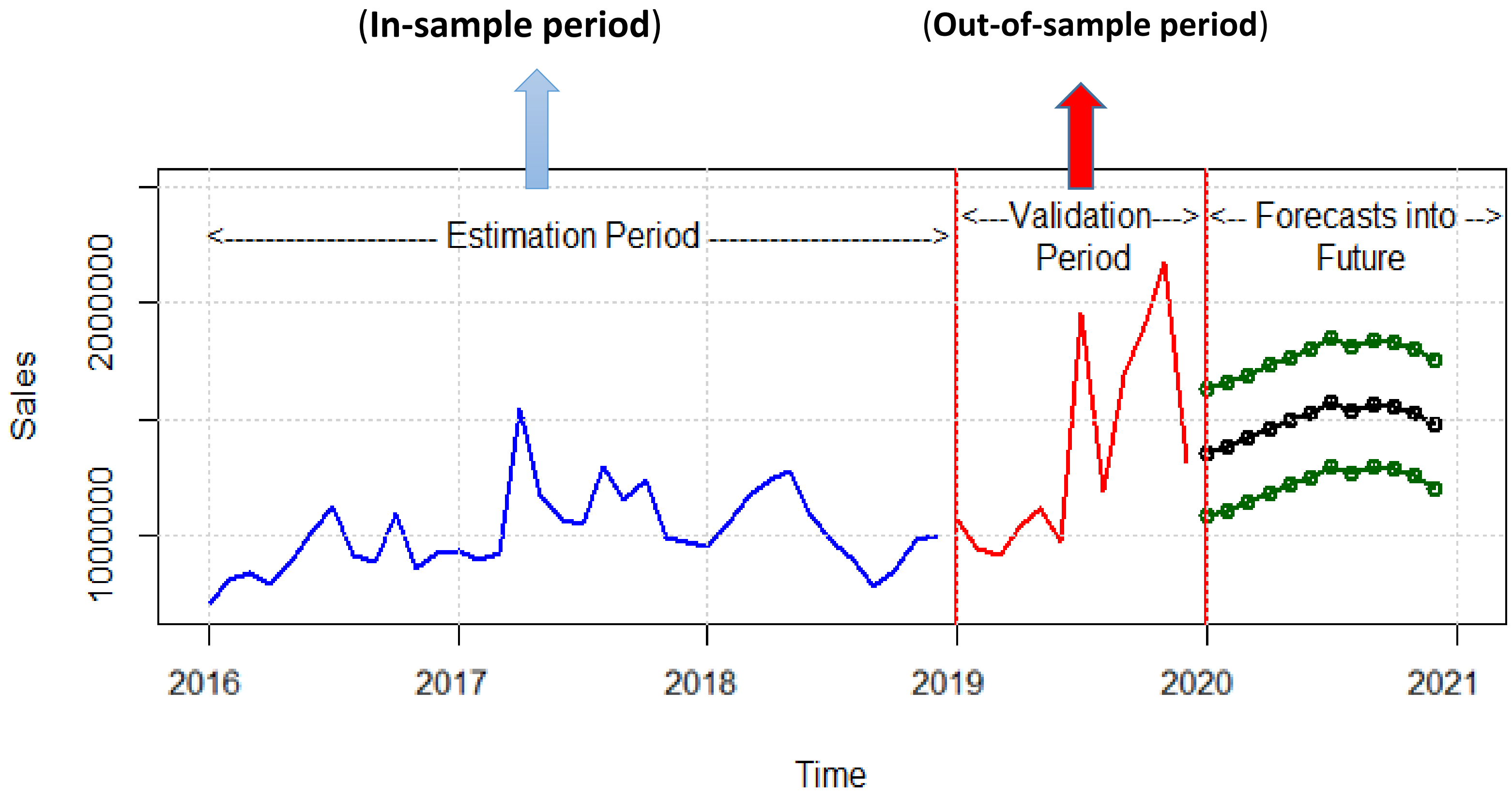
Other types of neural networks:

- Convolutional neural network – Good for image recognition
- Long short-term memory network – Good for speech recognition
- Recurrent neural network – Allows for temporal dynamic behaviour.



Forecast evaluation

Is the performance of my model good?



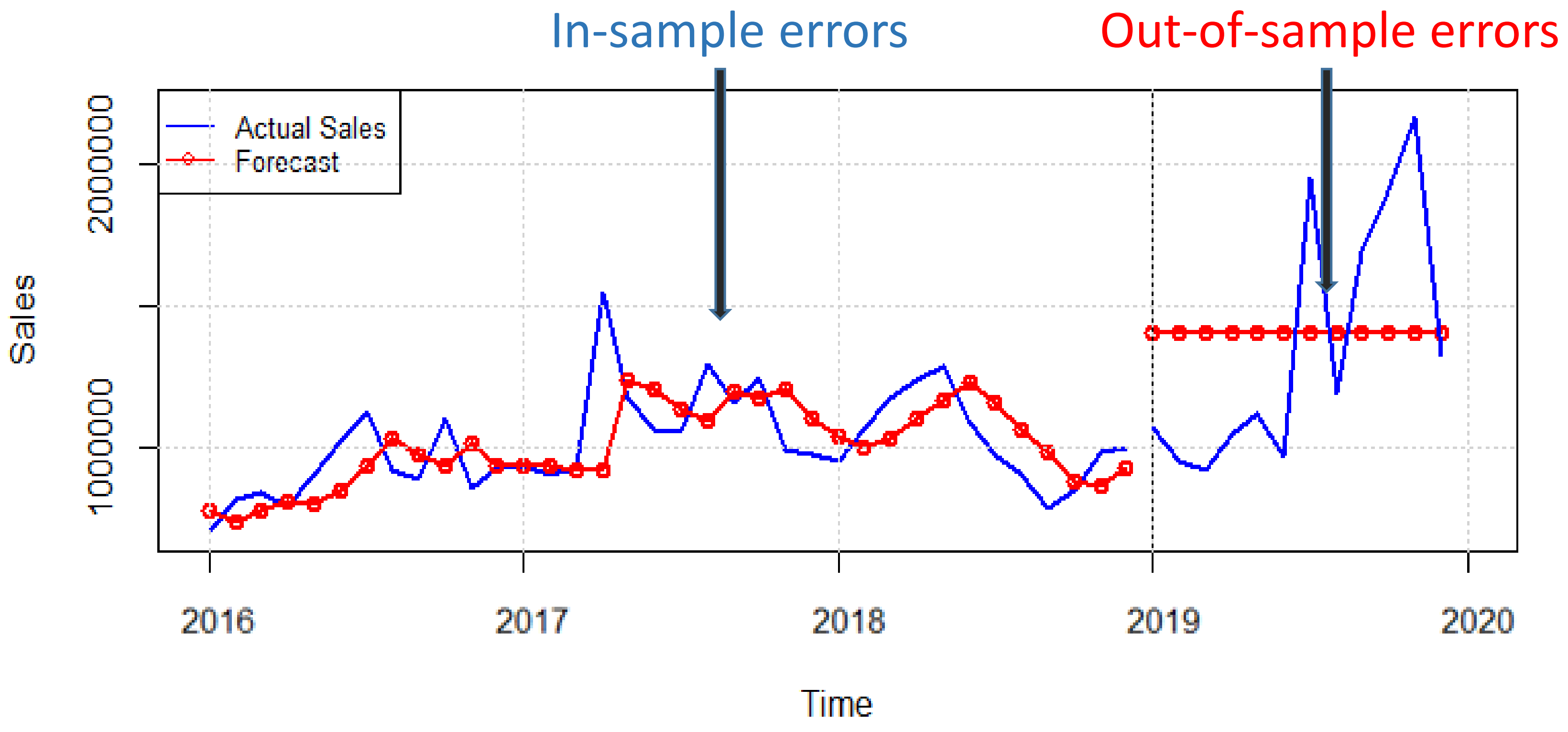
Forecast Errors

Error

Prediction

$$e_t = y_t - \hat{y}_t$$

Actual data



Forecast Error Measures

Mean Error (ME):

$$ME = \frac{1}{n}\sum_{t=1}^{n} e_t$$

Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |e_t|$$

Mean Squared Error (MSE):

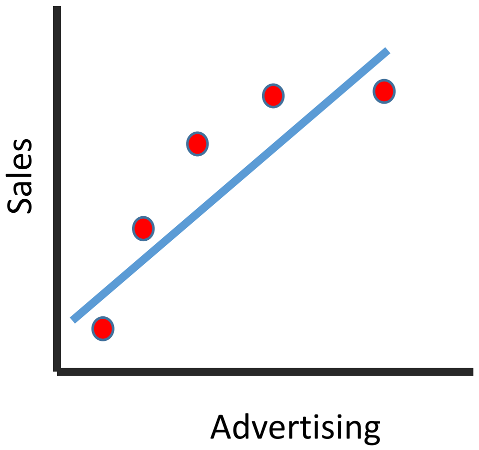
$$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2$$

Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \left|\frac{e_t}{y_t}\right| \times 100$$

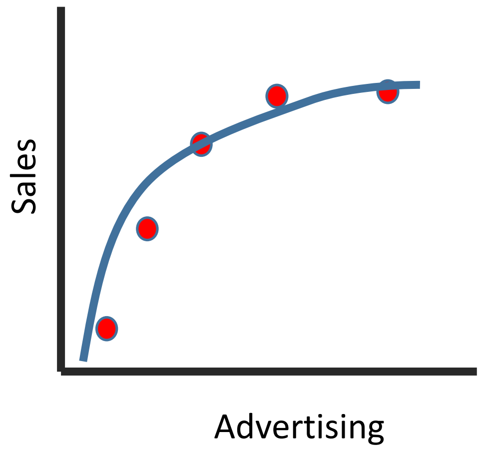
My training MSE is small, BUT…

1



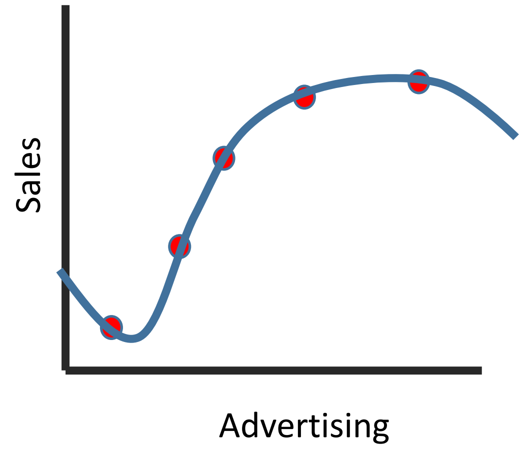
High bias
Low variance

2



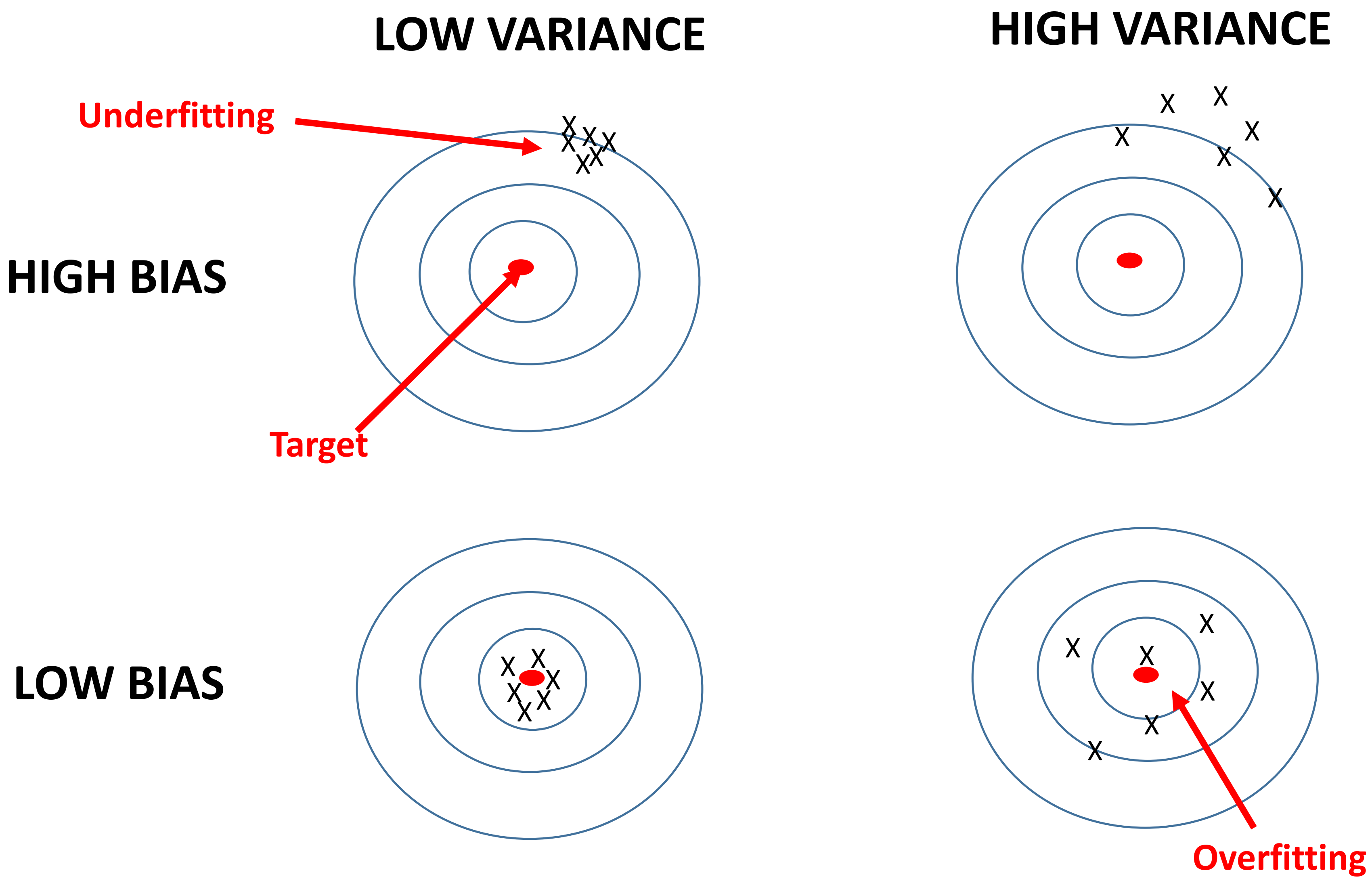
Moderate bias
Moderate variance

3



Low bias
High variance

Bias – Variance tradeoff



Bias – Variance tradeoff

In predictive modelling, there are three sources of uncertainty:

Estimation Error: The error in the coefficients when the linear model is true.

Model Bias: The error in the linear model when the true model is different (e.g. non-linear or one that contains other variables)

Irreducible Error: The noise in the data generating process, i.e. Data=Model + Noise.

$$\boxed{(Prediction\ Error)^2 = \sigma^2 + Bias^2 + Irreducible\ Error}$$

Validation

Validation techniques

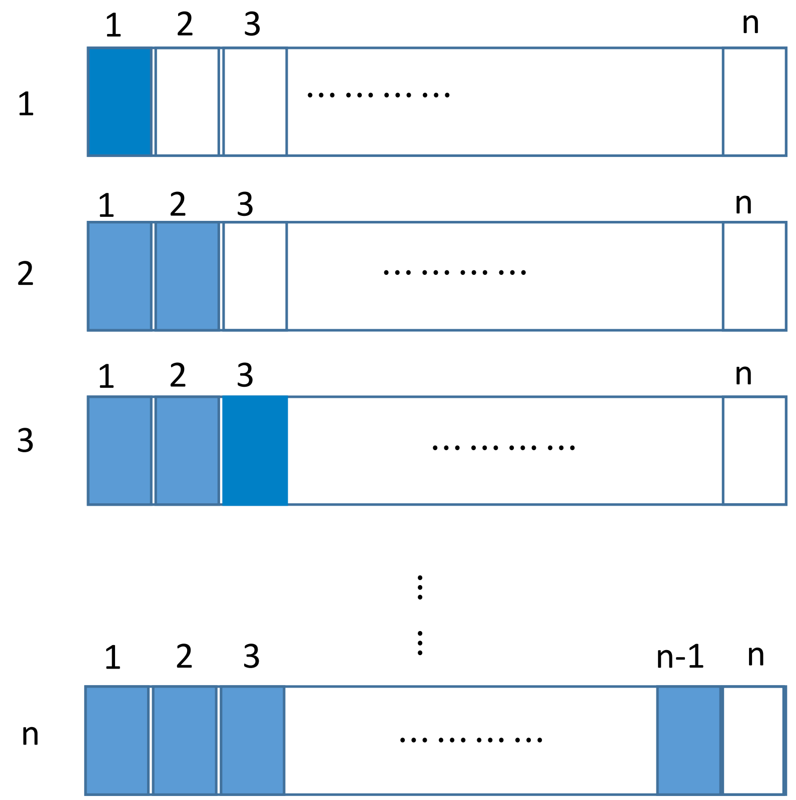
To overcome overfitting, pick the model that is good enough to fit the data without causing problems in the test set.

Use different training sets for validating your model and choose the model with the minimum error metric from the test data.

Cross Validation:

- Leave one out cross validation (LOOCV)
- k-fold cross validation
- Cross Validation for time series

Cross validation for time series



- Train fold 1 and test on *fold 2* through *fold n. Then, calculate the* MSE_1.
- Train using fold 1 and fold 2, and test on *fold 3* through *fold n. Then, calculate the* MSE_2.
- Train using fold 1, fold 2, and fold 3, and test on *fold 4* through *fold n. Then, calculate the* MSE_3.
- Train using *fold 1* through *fold n-1*, and test on *fold n. Then, calculate the* MSE_n.

Average MSE will be → $\frac{1}{n}\sum_{i=1}^{n} MSE_i$

Takeaways

Takeaways

- Importance of demand forecasting in retail
- Basic forecasting tools such as mean, naïve, SMA, Holt-Winters and ARIMA models.
- Forecasting in the presence of retail marketing interventions (e.g. price, promotions), using MLR and NNs.
- Forecast evaluation and validation for demand forecasts.

What is next?

- Demand forecasting with real-world datasets
- MLR and NNs
- Assignment (not compulsory)


ANY
QUESTIONS
?

Thank you!