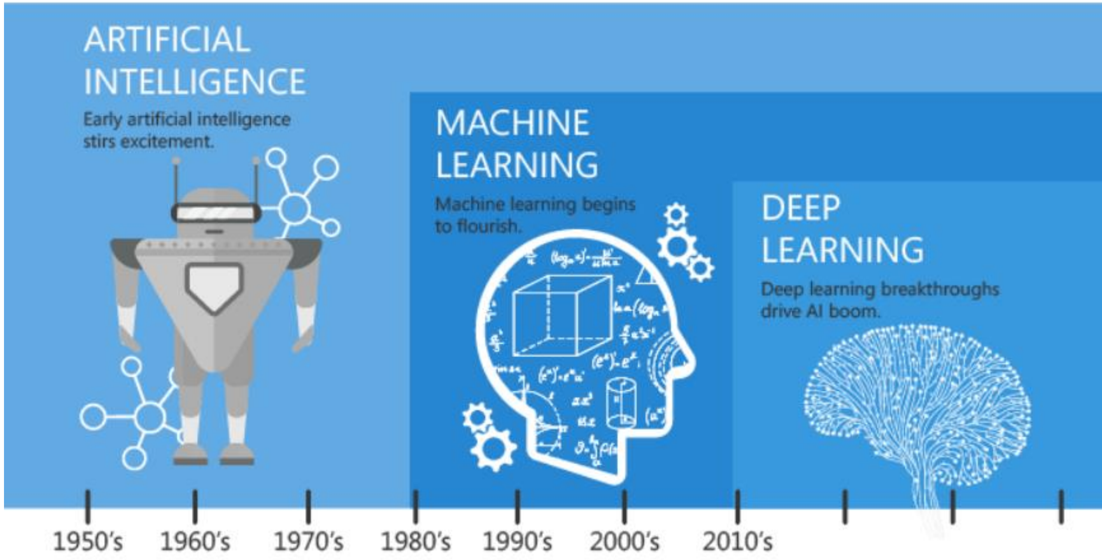
Lecture 17
Artificial Intelligence

Khola Naseem
khola.naseem@uet.edu.pk



Machine learning

➢Types of learning

➢ Machine learning implementations are classified into four major categories, depending on the nature of the learning

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning
4. Semi-supervised learning

Machine learning

- Issues: Data Preparation
  - Data cleaning
    - Preprocess data in order to reduce noise and handle missing values
  - Relevance analysis (feature selection)
    - Remove the irrelevant or redundant attributes
  - Data transformation
    - Normalize attribute values

Machine learning

- Classification Techniques:
  - Decision Tree based Methods
  - Rule-based Methods
  - Naïve Bayes
  - Neural Networks
  - K-nearest neighbor
  - SVM
  - and more...

Machine learning

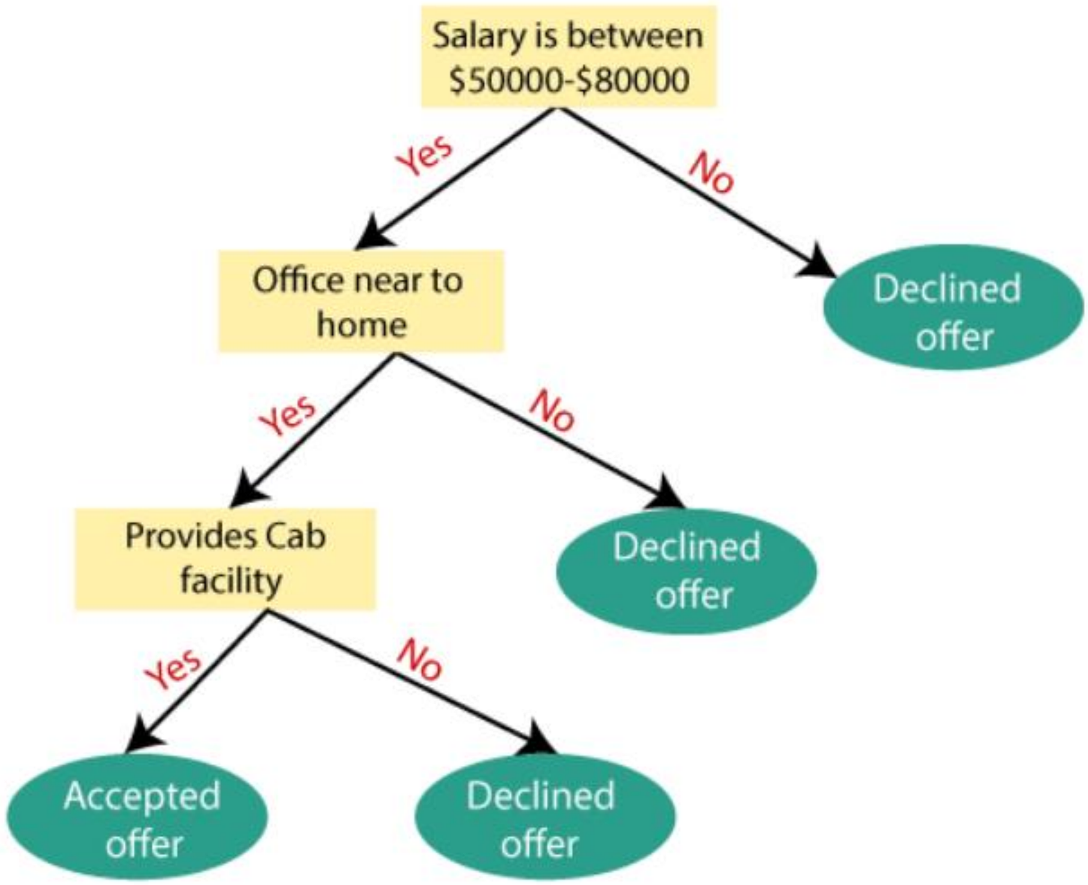
Classification Techniques: Decision Tree

- Decision Tree is a Supervised learning technique that mostly it is preferred for solving Classification problems.
- It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.
- It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.

Machine learning

➢Classification Techniques: Decision Tree

➢ Decision Tree

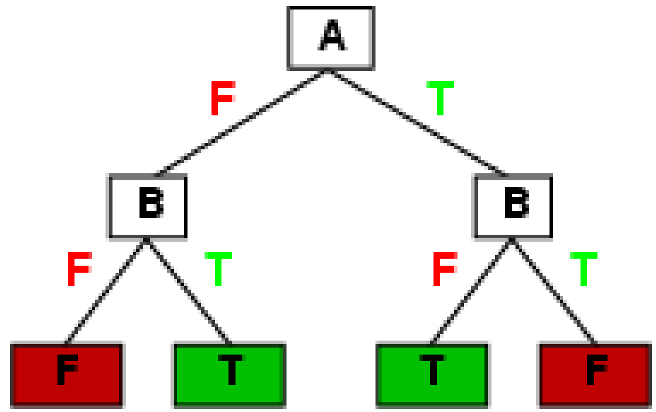


Credit: Khola Naseem

Machine learning

- Classification Techniques: Decision Tree based Methods
  - Decision trees can express any function of the input attributes.

A	B	A xor B
F	F	F
F	T	T
T	F	T
T	T	F

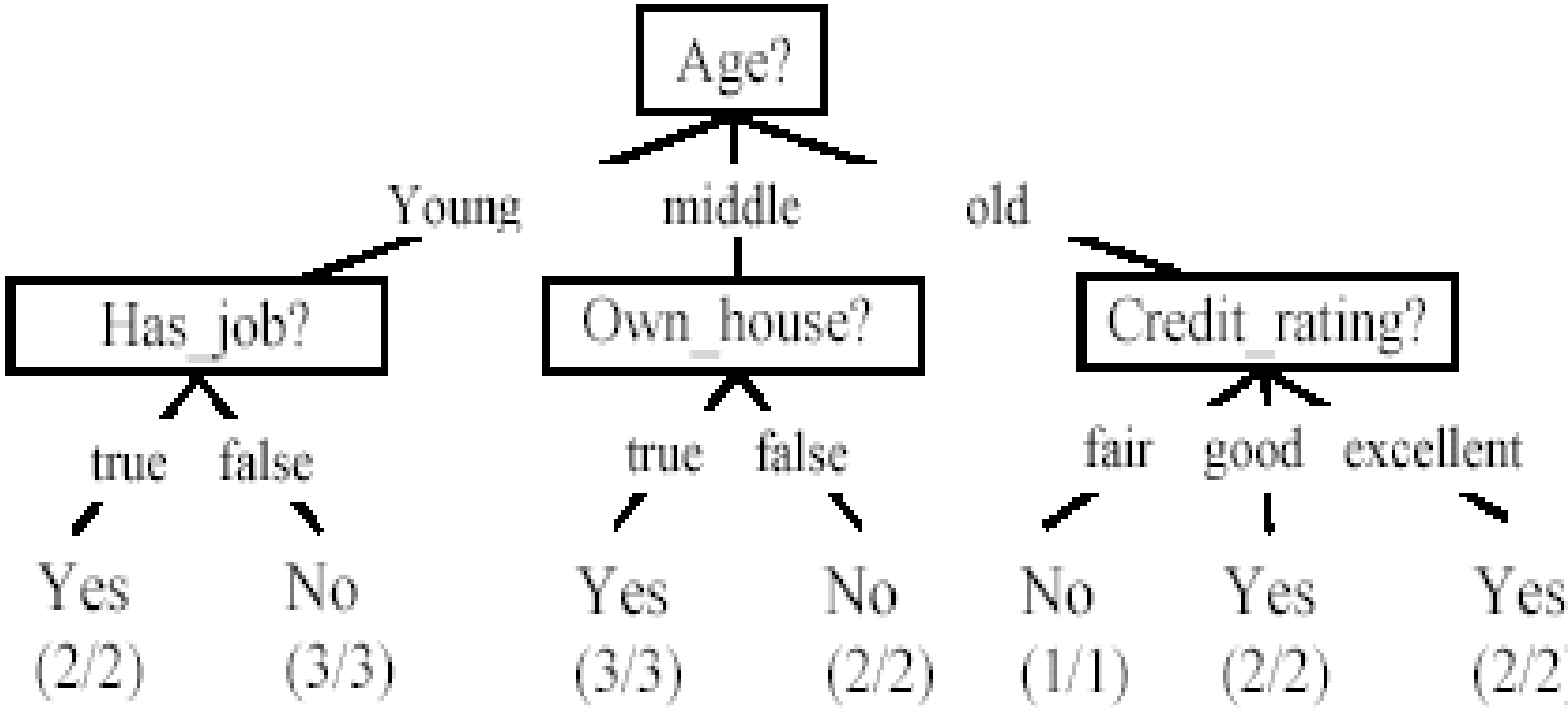


  - Trivially, there is a consistent decision tree for any training set with one path to leaf for each example (unless f nondeterministic in x) but it probably won't generalize to new examples

Credit: Khola Naseem

Machine learning

- Classification Techniques: Decision Tree based Decision nodes and leaf nodes (classes)



  - Problem

Credit: Khola Naseem

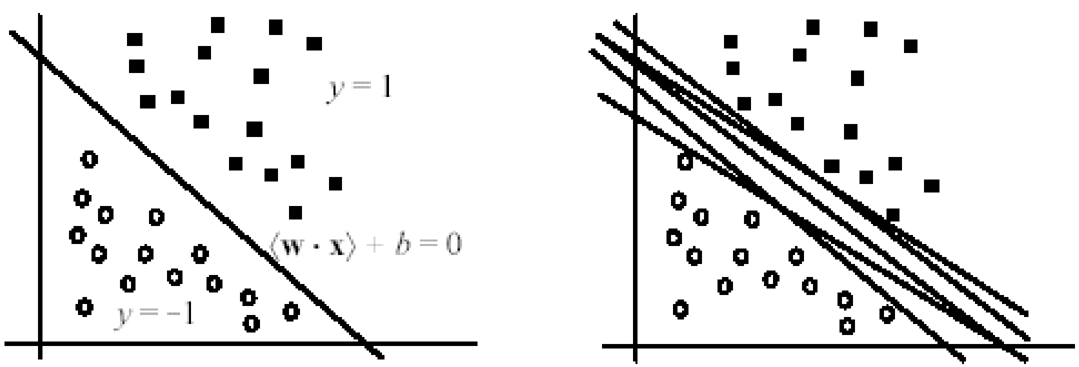
Machine learning

- Classification Techniques: Rule based
  - R1: (Give Birth = no) ^ (Can Fly = yes) => Birds
  - R2: (Give Birth = no) ^ (Live in Water = yes) => Fishes
  - R3: (Give Birth = yes)^ (Blood Type = warm) => Mammals
  - R4: (Give Birth = no) ^ (Can Fly = no) => Reptiles
  - R5: (Live in Water = sometimes) => Amphibians

Machine learning

- Classification Techniques: SVM
  - SVMs are **linear classifiers** that find a hyperplane to separate **two class** of data, positive and negative.
  - Kernal to handle non linearity
  - $f(\mathbf{x}) = \langle \mathbf{w} \cdot \mathbf{x} \rangle + b$

$$y_i = \begin{cases} 1 & if \ \langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b \geq 0 \\ -1 & if \ \langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b < 0 \end{cases}$$



Credit: Khola Naseem

Machine learning

- Classification Techniques: k-Nearest Neighbor Classification (kNN)
- To classify a test instance *d*, define *k*-neighborhood *P* as *k* nearest neighbors of *d*

Algorithm kNN(D, d, k)

Compute the distance between d and every example in D;
Choose the k examples in D that are nearest to d, denote the set by P ($\subseteq D$);
Assign d the class that is the most frequent class in P (or the majority class);

- *Distance function is crucial, but depends on applications.*

Credit: Khola Naseem

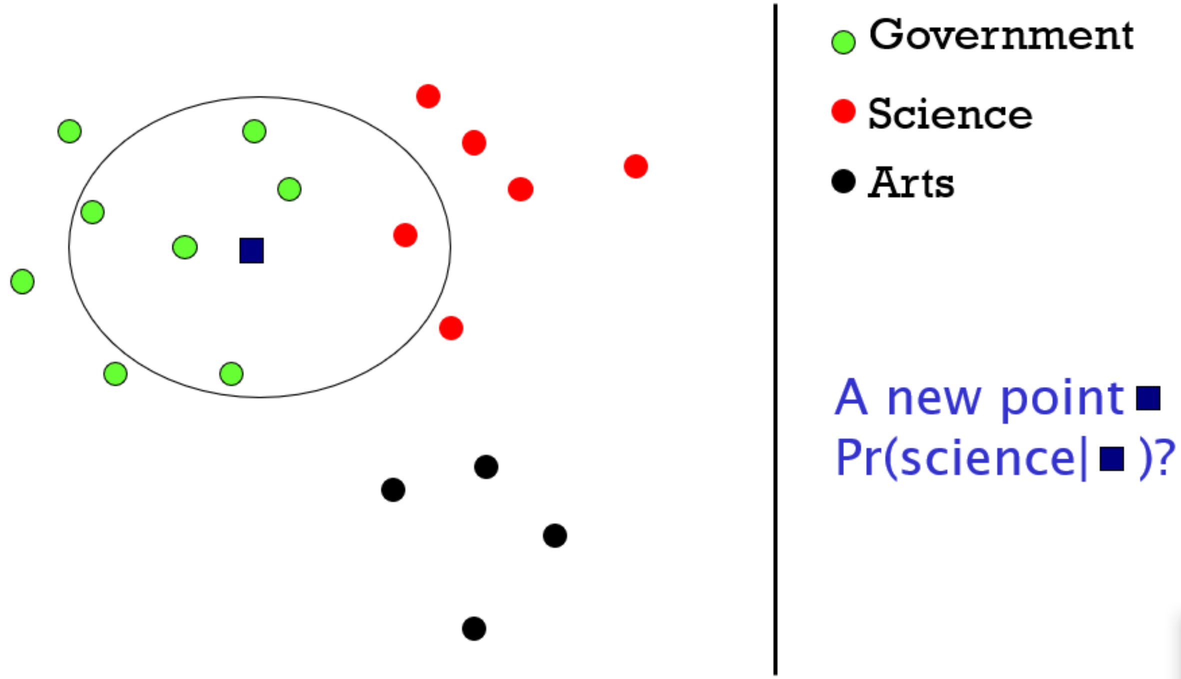
Machine learning

- Classification Techniques: k-Nearest Neighbor Classification (kNN)

Algorithm kNN(D, d, k)

1 Compute the distance between d and every example in D;
2 Choose the k examples in D that are nearest to d, denote the set by P ($\subseteq D$);
3 Assign d the class that is the most frequent class in P (or the majority class);

- K=6



Machine learning

- Classification Techniques:
  - This large number of methods also show the importance of classification and its wide applicability.
  - It remains to be an active research area.

Machine learning

- Model Evaluation Metrics for Performance Evaluation:
  - How to evaluate the performance of a model?
  - Methods for Performance Evaluation
  - How to obtain reliable estimates?
  - Methods for Model Comparison
  - How to compare the relative performance of different models?

Machine learning

- Evaluation metrics:
- Focus on the predictive capability of a model Rather than how fast it takes to classify or build models, scalability, etc.
- Confusion Matrix:

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Credit: Khola Naseem

Machine learning

- **Evaluation metrics:**

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

- **Accuracy**

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

- **Limitation of Accuracy**
  - consider a 2-class problem
  - Number of Class 0 examples = 9990
  - Number of Class 1 examples = 10
  - If model predicts everything to be class 0,
  - accuracy is 9990/10000 = 99.9 %
  - Accuracy is misleading because model does not detect any class 1 example

Credit: Khola Naseem

Machine learning

- **Evaluation metrics:**
- **Precision and recall**
- Suppose that $y = 1$ in presence of a rare class that we want to detect
- **Precision** (How much we are precise in the detection) Of all patients where we classified $y = 1$, what fraction actually has the disease?

$$\frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

Machine learning

- **Evaluation metrics:**
- **Precision and recall**
- Suppose that $y = 1$ in presence of a rare class that we want to detect
- **Recall** (How much we are good at detecting)
- Of all patients that actually have the disease, what fraction did we correctly detect as having the disease?

$$\frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

Machine learning

- **Evaluation metrics:**

- **F1-score**

- It is usually better to compare models by means of one number only. The **F1 – score** can be used to combine precision and recall

$$\frac{2TP}{2TP + FP + FN}$$

Machine learning

- **Evaluation metrics:**

	1 (p)	0 (n)
1 (Y)	True positive 11	False positive 0
0 (N)	False negative 1	True negative 11