Anomaly Detection & Anomaly Detection Approaches

Unit objectives

After completing this unit, you should be able to:

- Understand the applications of anomaly detection
- Learn about example of classification-based methods
- Gain knowledge on nearest neighbor-based approach
- Gain an insight into clustering based methods
- Understand statistical approach, graph and model-based approach for ML implementation

What are anomalies?

- The detection of anomalies is a method used to detect unusual patterns not in accordance with appropriate behavior, called outliers.

- This includes many business applications:
  - Intrusion detection(recognizing unusual anomalies in network traffic activity that may indicate a hack).
  - System health monitoring (recognizing a cancerous tumor growing in an MRI scan).
  - Fraud detection in transactions with credit card to failure detection in operational environments.

Applications of anomaly detection

- Incursion/Intrusion discovery.
- Monitoring of frauds/ crimes.
- Healthcare informatics.
- Detection of industrial damage.
- Image processing.

Related use cases

- Rare class mining.
- Chance discovery.
- Novelty detection.
- Exception mining.
- Removal of noise.

Types of input data

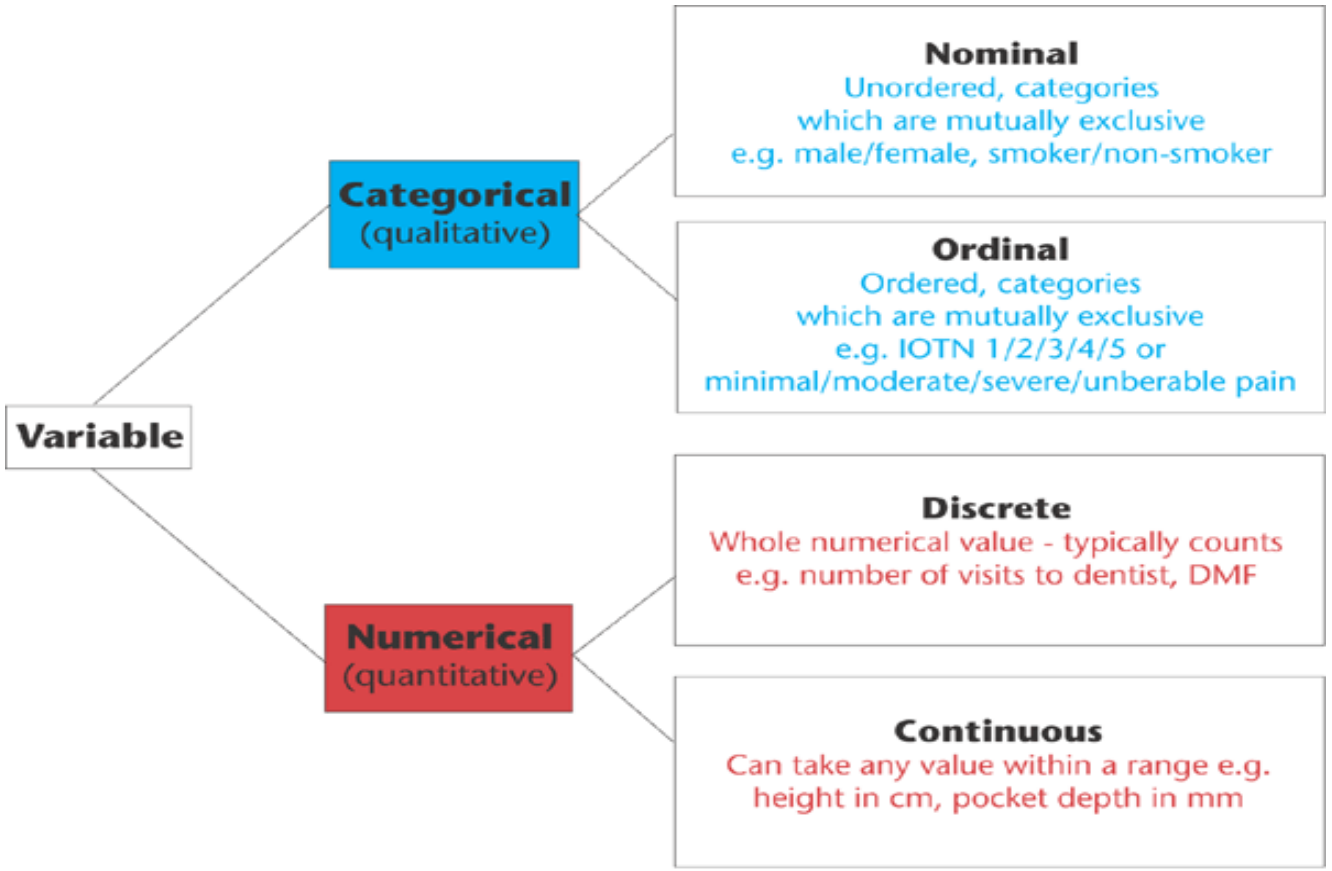

Figure: Types of Input Data

Source: https://images.app.goo.gl/r1NtrYL4DiHSu2y69

Types of anomalies

- Anomalies can be classified into the following three categories:
  - Point anomalies:
    - If one component is aberration, it is an anomaly purpose against all the other items.
    - It is the easiest type of phenomenon and it is studied by other studies.
    - The variations of points are considered in O1 and O2.
  - Contextual anomalies:
    - If in any given sense the entity is aberrant. just here is it a conceptual abnormality (also known as conditional anomaly).
  - Collective anomalies:
    - If irregularity is seen with a few other objects connected to other particles.
    - Such scenario, even the set of artifacts should not be aberrant.

Evaluation of an anomaly detector

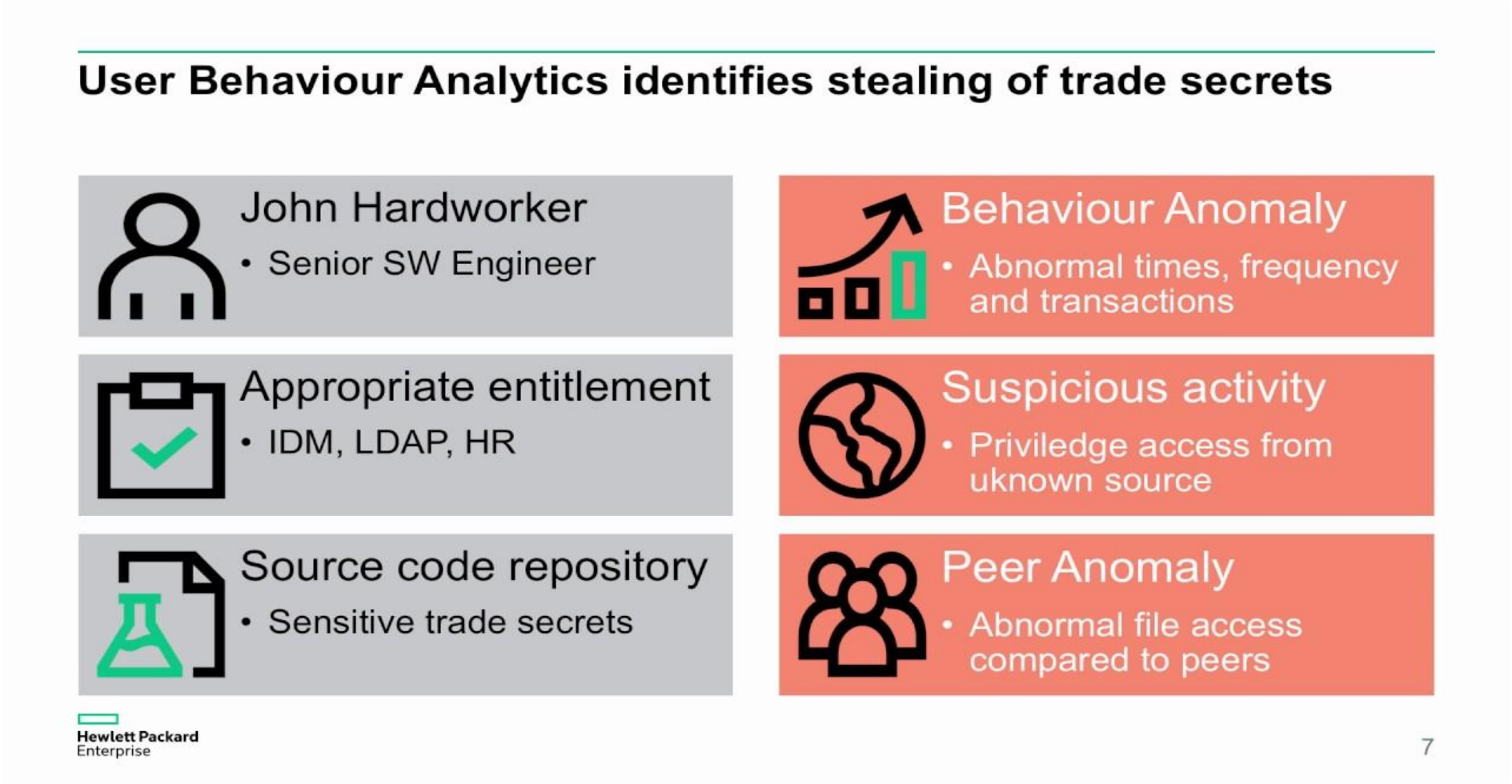

Figure: UBA Analysis

Source: https://images.app.goo.gl/S4XWcMZNQKmTec2Z7

Taxonomy of approaches

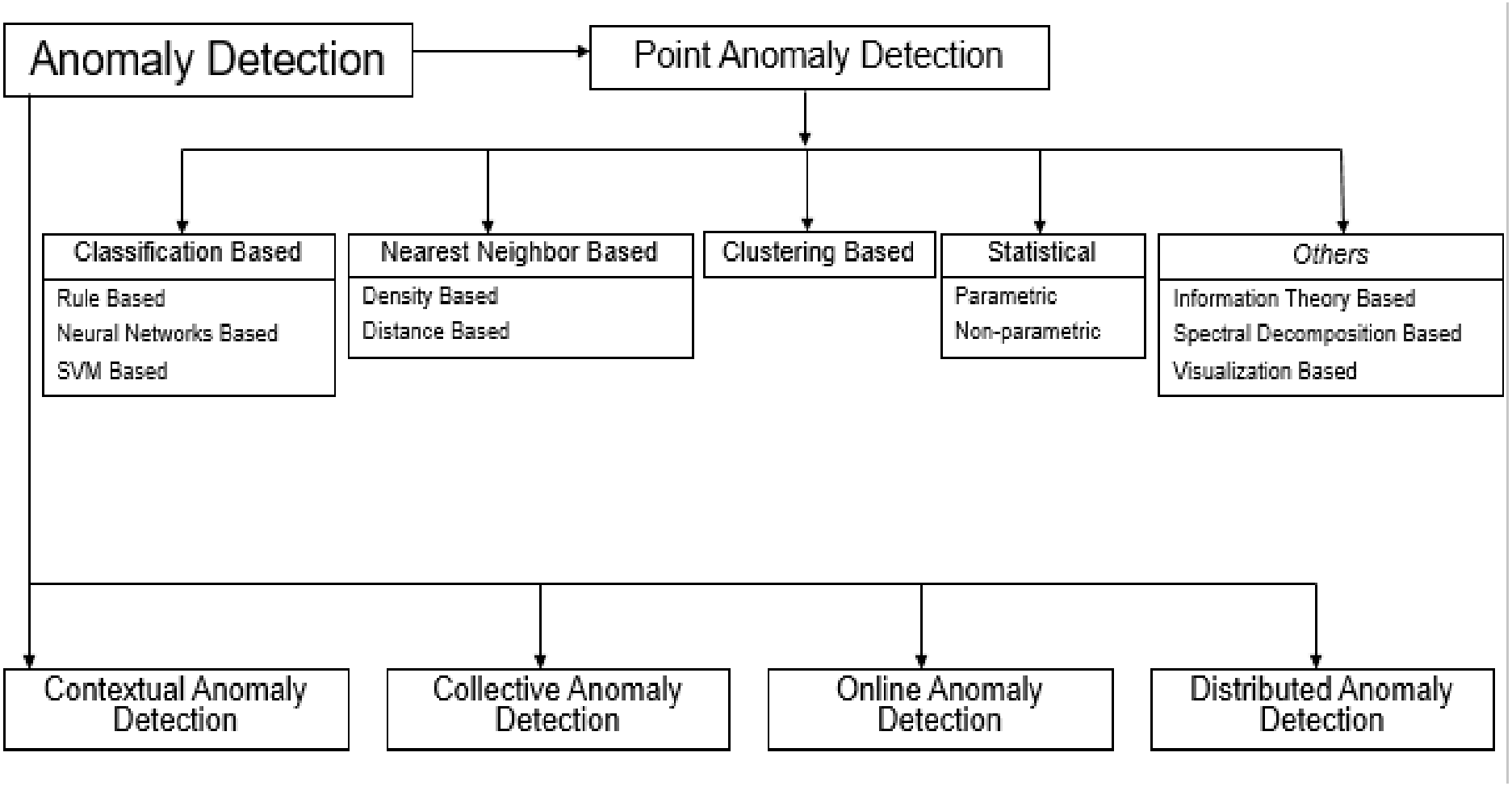

Figure: Approaches of taxonomy

Classification based

- The primary reason: Create a categorization models based on defined dataset for usual (and outlier (uncommon) occurrences to identify any mysterious new incident.

- Slanted (extremely unbalanced) class distributions will be treated by classification methods.

- Classification:
  - Methods for controlled sorting.
  - Need regular category & aberrant type understanding.
  - Define the structure to distinguish natural from established abnormalities.

- Strategies of semi-controlled grouping:
  - Need ordinary class awareness only.
  - Use the adapted detection data to predict behavior pattern and then recognize any abnormalities behavior as unusual.

Classification use cases

- Interception of web traffic-certain.
- Video classification.
- Classification of photographs.
- Classification of speech.

Supervised classification techniques

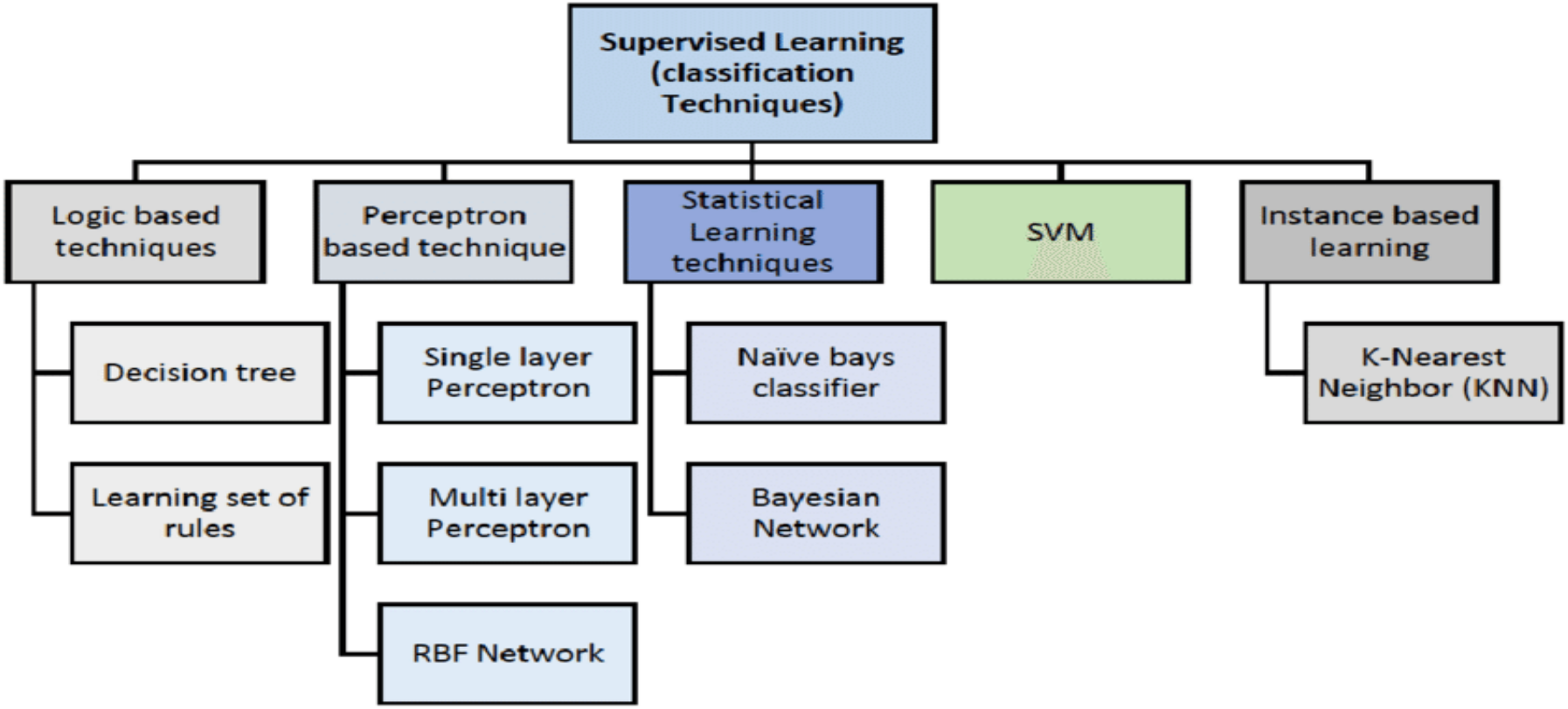

Figure: Supervised Classification Techniques

Source: https://images.app.goo.gl/PyZtjRJLce2FDQ6i7

Self evaluation: Exercise 14

- To continue with the training, after learning the various steps involved in pattern recognition and anomaly detection, it is instructed to utilize the concepts to perform the following activity.

- You are instructed to write the following activities using python code.

- Exercise 14: Cluster based Local Outlier Factor (CBLOF).

Nearest neighbor based techniques

- Relevant presumption:
  - Standard points are in proximity, and anomalies are far away from other points.
  - 2 step common method:
    - For every data record, compute community.
    - To assess if or not the data reports are anomalous in the community.

- Categories:
  - Distance based methods: Anomalies are by far the most isolated reference points.
  - Density based methods: Data points in low-density regions are exceptions.

Self evaluation: Exercise 15

- To continue with the training, after learning the various steps involved in pattern recognition and anomaly detection, it is instructed to utilize the concepts to perform the following activity.

- You are instructed to write the following activities using python code.

- Exercise 15: Local Density Cluster based Outlier Factor (LDCOF).

Self evaluation: Exercise 16

- To continue with the training, after learning the various steps involved in pattern recognition and anomaly detection, it is instructed to utilize the concepts to perform the following activity.

- You are instructed to write the following activities using python code.

- Exercise 16: Local Correlation Integral (LOCI).

Others model techniques

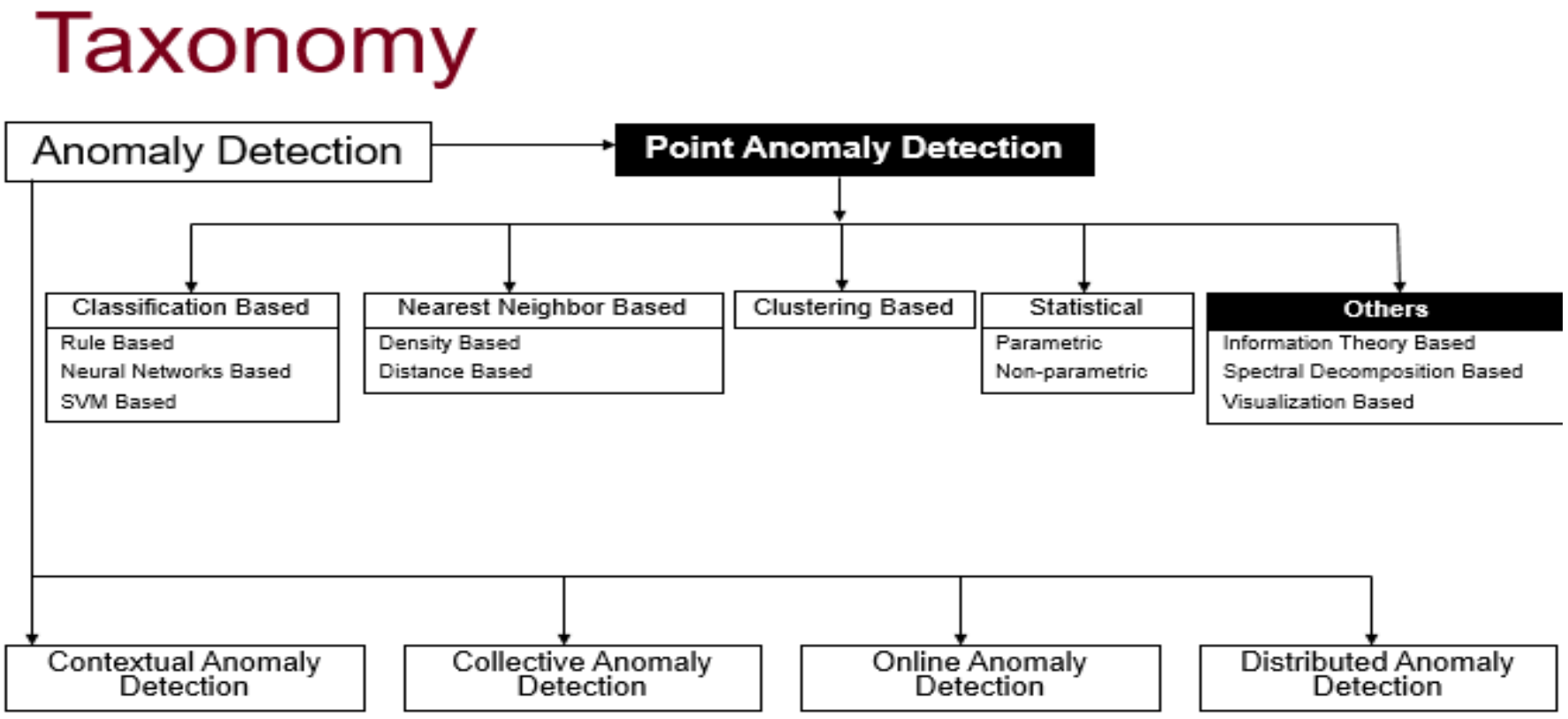

Figure: Others model techniques

Information theory

- In information theory, the main purpose is for one person (a transmitter) to send a message to another (the recipient) over a path.

- To do that, the transmitter sends a sequence of partial messages (possibly one) that provide hints to the original message.

- The details value in each of these incomplete communications is an indicator of how ambiguous it is for the receiver. For starters:
  - A partial message which reduces in half the number of possibilities transmits a bit of message details.
  - If, for example, the transmitter wanted to submit the output of a randomly chosen digit to the recipient, the partial response of "the number is odd" would give a bit of details.

Contextual anomaly based

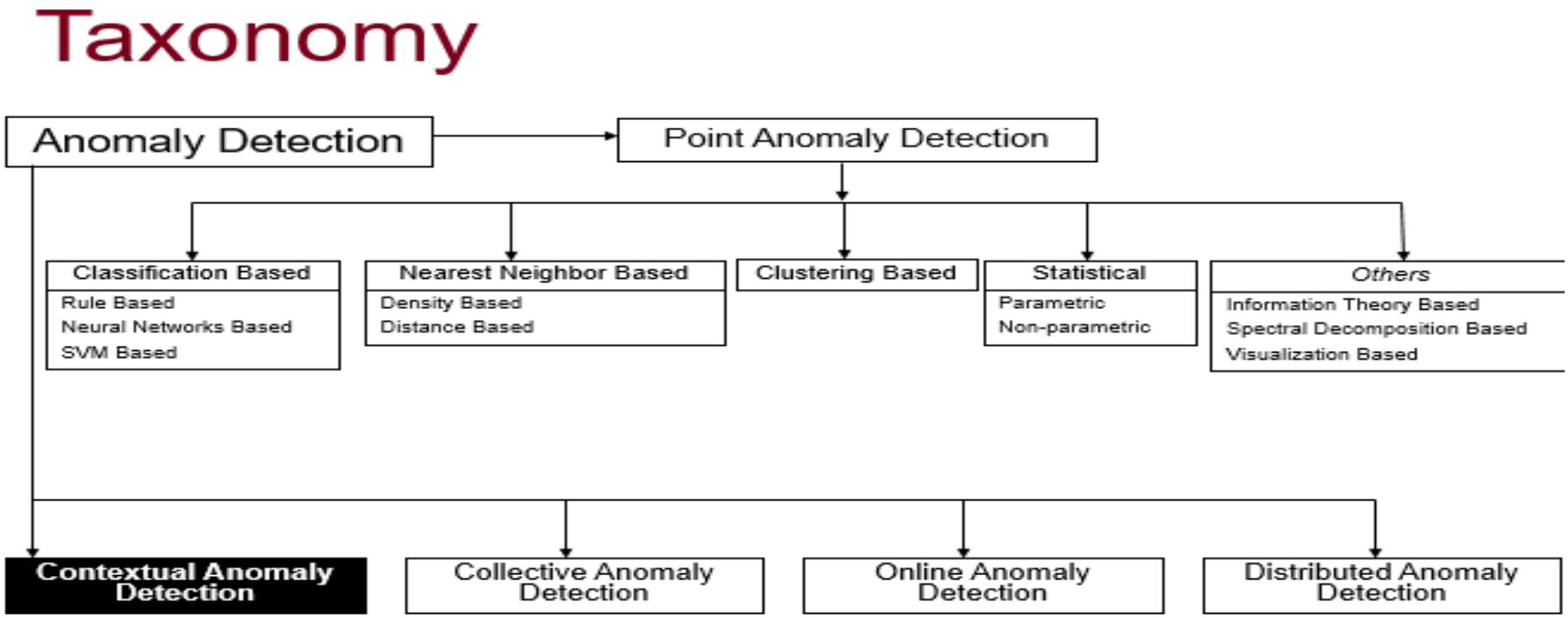

Figure: Contextual anomaly

Self evaluation: Exercise 17

- To continue with the training, after learning the various steps involved in pattern recognition and anomaly detection, it is instructed to utilize the concepts to perform the following activity.

- You are instructed to write the following activities using python code.

- Exercise 17: Influenced Outlierness (INFLO).

Collective anomaly detection

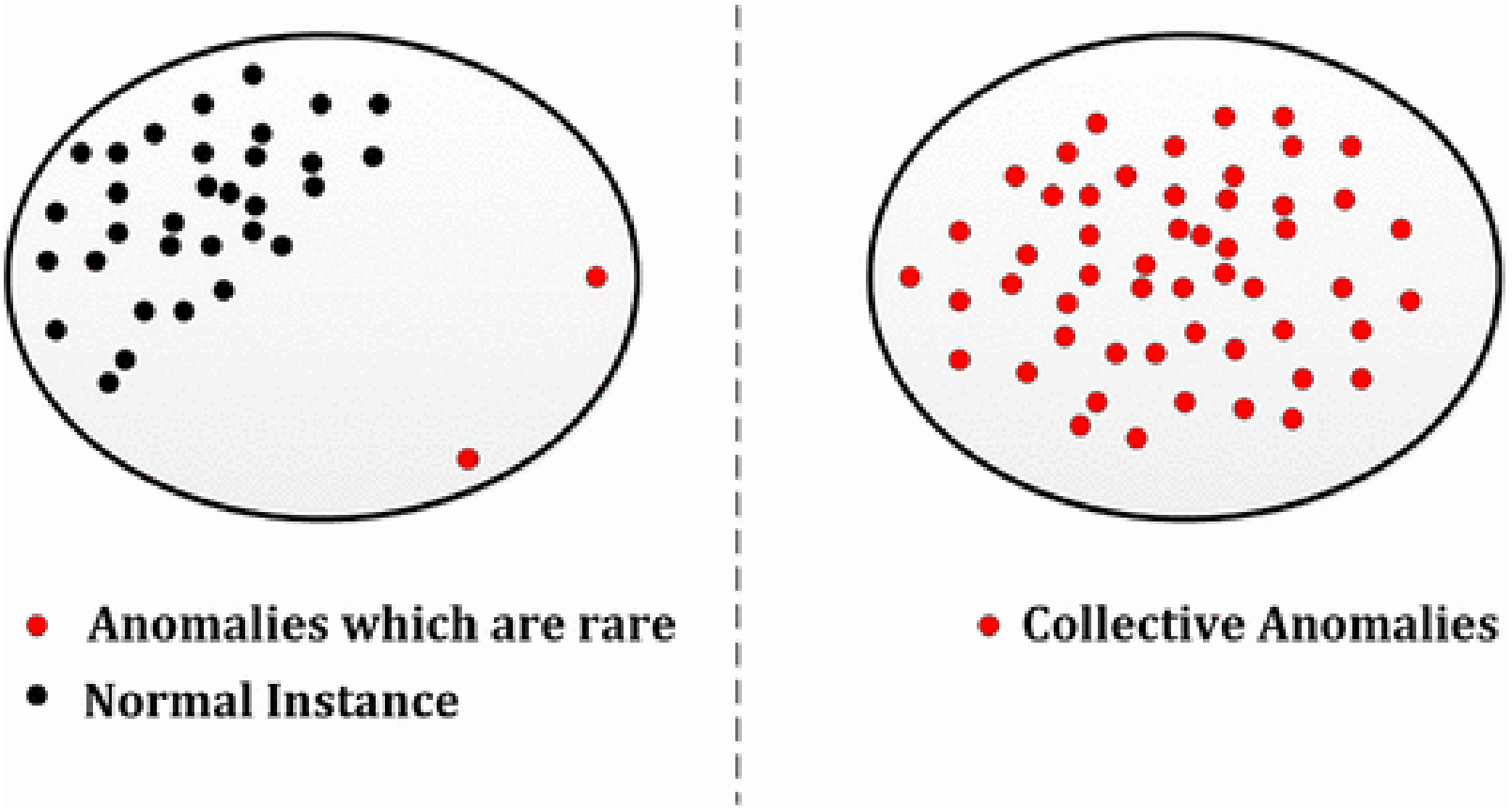

Figure: Collective Anomaly Detection

Source: https://images.app.goo.gl/Uw64MLJ8k1ewrun37

On-line based model

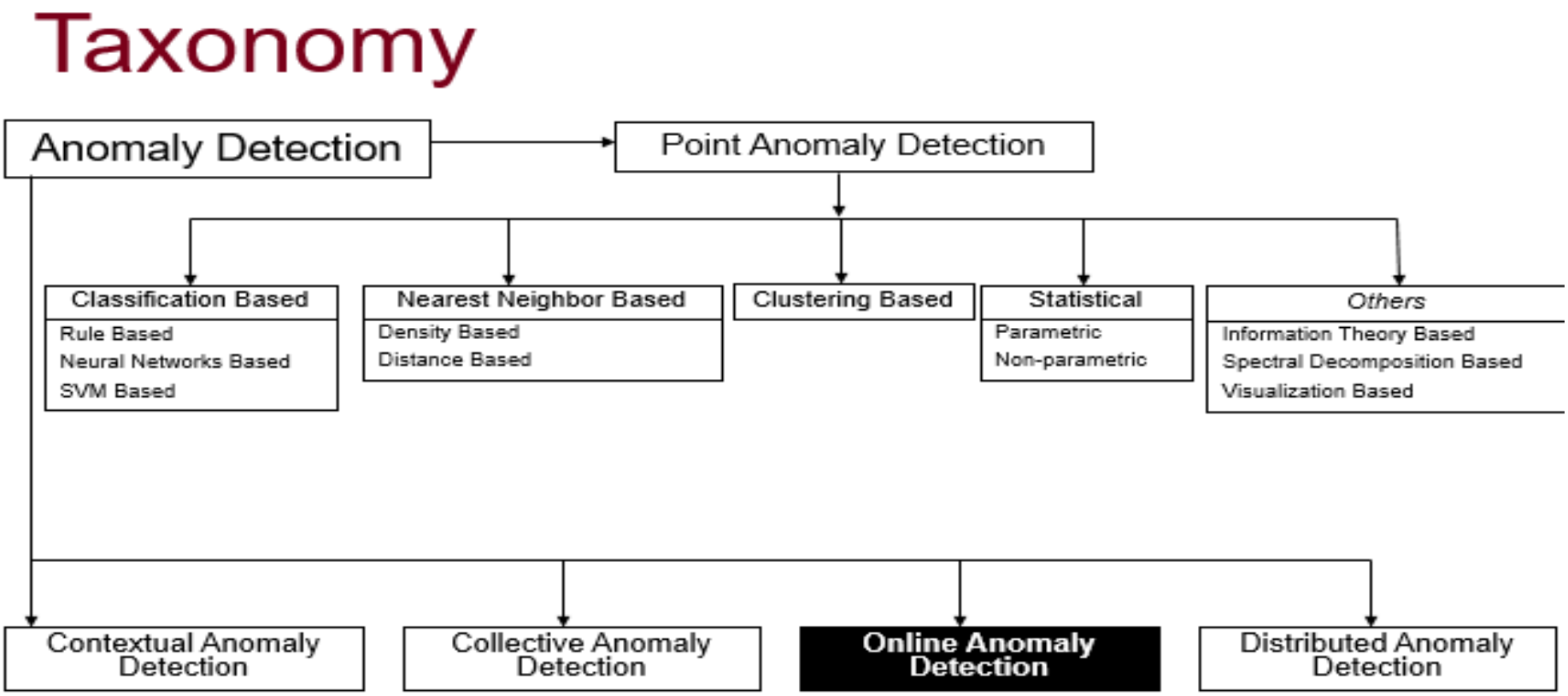

Figure: On-line based model

Distributed anomaly detection

- Data could originate from various sources in several phenomenon object tracking:
  - Intrusion prevention network.
  - Misuse by payment card.
  - Health of aircraft.

- Examination of information from a data position will un-detect errors that take place at many points concurrently:
  - In such hierarchical structures, abnormalities can be observed by combining information from predefined intervals on identified abnormalities in order to identify irregularities at national level in a complicated web.

- Good efficiency and decentralized architectures are required for connection and anomaly incorporation.

Self evaluation: Exercise 18

- To continue with the training, after learning the various steps involved in pattern recognition and anomaly detection, it is instructed to utilize the concepts to perform the following activity.

- You are instructed to write the following activities using python code.

- Exercise 18: Local Outlier Probability (LoOP).

IDS analysis strategy

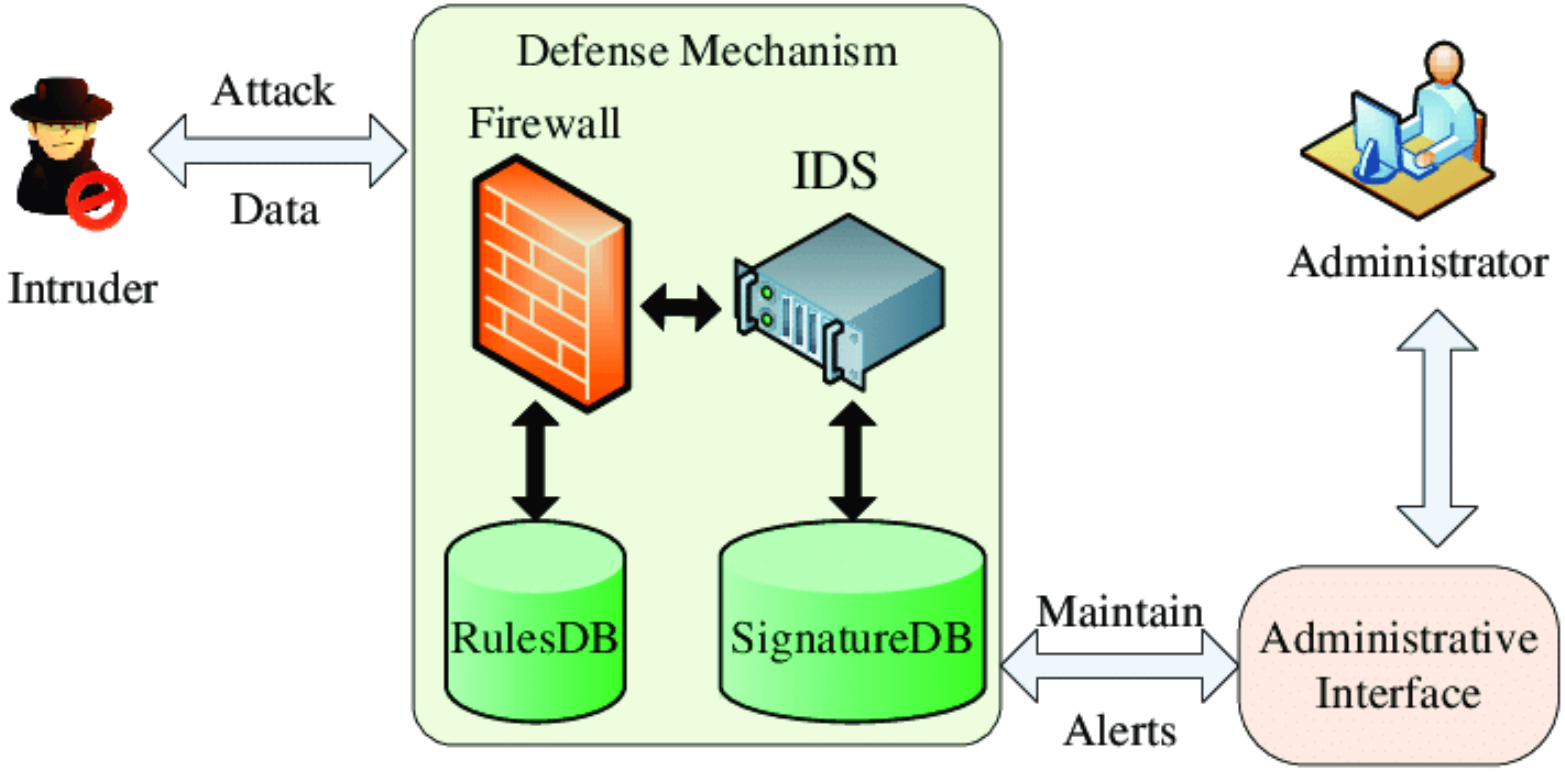

Figure: IDS

Source: https://images.app.goo.gl/gzuRHTQH2FAVeuVN8

Self evaluation: Exercise 19

- To continue with the training, after learning the various steps involved in pattern recognition and anomaly detection, it is instructed to utilize the concepts to perform the following activity.

- You are instructed to write the following activities using python code.

- Exercise 19: Connectivity based Outlier Factor (COF).

Checkpoint (1 of 2)

Multiple choice questions:

1. What is unsupervised learning?
   a) Features of group explicitly stated
   b) Number of groups may be known
   c) Neither feature & nor number of groups is known
   d) None of the mentioned

2. What is plasticity in neural networks?
   a) Input pattern keeps on changing
   b) Input pattern has become static
   c) Output pattern keeps on changing
   d) Output is static

3. What are the tasks that cannot be realized or recognized by simple networks?
   a) Handwritten characters
   b) Speech sequences
   c) Image sequences
   d) All of the mentioned

Checkpoint solutions (1 of 2)

Multiple choice questions:

1. What is unsupervised learning?
   a) Features of group explicitly stated
   b) Number of groups may be known
   c) Neither feature & nor number of groups is known
   d) None of the mentioned

2. What is plasticity in neural networks?
   a) Input pattern keeps on changing
   b) Input pattern has become static
   c) Output pattern keeps on changing
   d) Output is static

3. What are the tasks that cannot be realized or recognized by simple networks?
   a) Handwritten characters
   b) Speech sequences
   c) Image sequences
   d) All of the mentioned

Checkpoint (2 of 2)

Fill in the blanks:

1. Expectation maximization is an algorithm ____ in machine learning.
2. The only examples necessary to compute ____in an SVM is support vectors.
3. Averaging out the predictions of multiple _____will drastically reduce the variance.
4. The presence of _____(which leads to overfitting) is not generally a problem with weak classifiers.

True or False:

1. MAP estimates are equivalent to the ML estimates when the prior used in the MAP is a uniform prior over the parameter space. True/False
2. Because decision trees learn to classify discrete-valued outputs instead of real-valued functions it is impossible for them to over fit. True/False
3. If $P(A|B) = P(A)$ then $P(A \cap B) = P(A)P(B)$. True/False

Checkpoint solutions (2 of 2)

Fill in the blanks:

1. Expectation maximization is a **clustering** algorithm in machine learning.
2. The only examples necessary to compute **f(x)** in an SVM is support vectors.
3. Averaging out the predictions of multiple **classifiers** will drastically reduce the variance.
4. The presence of **over-training** (which leads to overfitting) is not generally a problem with weak classifiers.

True or False:

1. MAP estimates are equivalent to the ML estimates when the prior used in the MAP is a uniform prior over the parameter space. **True**
2. Because decision trees learn to classify discrete-valued outputs instead of real-valued functions it is impossible for them to over fit. **False**
3. If $P(A|B) = P(A)$ then $P(A \cap B) = P(A)P(B)$. **True**

Question bank

Two marks question:

1. How would you handle an imbalanced dataset?
2. What's the F1 score? How would you use it?
3. Which is more important to you model accuracy, or model performance?
4. How is a decision tree pruned?

Four marks question:

1. How would you handle an imbalanced dataset?
2. When should you use classification over regression?
3. Name an example where ensemble techniques might be useful.
4. How do you ensure you're not overfitting with a model?

Eight marks question:

1. How would you evaluate a logistic regression model?
2. What's the "kernel trick" and how is it useful?

Unit summary

Having completed this unit, you should be able to:

- Understand the applications of anomaly detection
- Learn about example of classification-based methods
- Gain knowledge on nearest neighbor-based approach
- Gain an insight into clustering based methods
- Understand statistical approach, graph and model-based approach for ML implementation