**The topics for the examination**

Definition of Data Mining

Examples of Data Mining tasks

Challenges in Data Mining

Attributes and Objects in Data Mining

What is data

Types of Attributes (Nominal, Ordinal, Interval, Ratio)

Properties of Attribute Values

Discrete and Continuous Attributes

Asymmetric Attributes

Data Quality problems ( Noise, Outliers)

Types of Datasets

Missing Values (MCAR, MAR, MNAR etc)

Different types of distances (Euclidean, Manhattan, Minkowski)

What is a metric

Visually evaluating correlation

Drawback of correlation

Meaning of Information in information theory

Entropy

Mutual Information

Data Preprocessing

Aggregation

Sampling (types of sampling)

Dimensionality reduction, curse of dimensionality

Feature subset selection

Feature Creation

Discretization and binarization

Attribute transformation

Important Characteristics of Data

Decision Trees

Accuracy

Error Rate

Entropy

Information Gain

Gini Index

Decision Boundary

Model Overfitting, Underfitting

Reasons for Model Overfitting

Occam’s Razor

Train set, test set and Validation set

Cross validation

Perceptron

Perceptron weight update rule