Descriptive Learning Methods (Unsupervised)

K-means, moving centers

Market Basket Analysis

Kohonen Network (SOM)

CPA, CMA, CA

Cluster Analysis

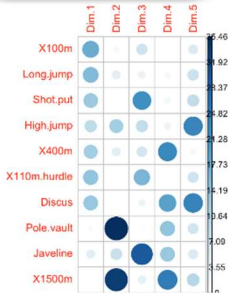
Clusters similar data points to each other that are less like those in separate clusters.

E.g. Consider a bookstore with different areas for the types of books: History, Self-Help, Romance, Mystery & Crime etc. The books in each of these clusters are more like each other than they are to other clusters.

PCA

Help you identify which variables are important so you can compress the data by reducing the number of dimensions.

E.g. gender, days of week they typically shop, amount spent on average trip.



Association Rules/Analysis

In a given set of records, each will contain a number of items. Association analysis allows you to determine the degree to which the items tend to be associated with one another.

E.g. people who buy hamburger buns will also likely buy mustard and hamburger meat. You can associate the items together and create rules.

Neural Clustering

Classification method - AKA Self-Organizing Map (SOM)

Unsupervised learning **neural network**

In the end of the learning, all inputs are classified as the node they are closest to. Each input that is processed changes the location of the nodes.

No variables to predict, it learns the structure of the data in order to distinguish clusters hidden within the data.

Operates like a matrix that is made up of cells (nodes), vectors, and magnitudes (weights) of each vector.

Factor Analysis

Multiple observed variables have similar patterns of responses because they are all associated with a latent *(i.e. not directly measured)* variable.

Involves grouping variables with similar attributes into a matrix using linear algebra techniques.

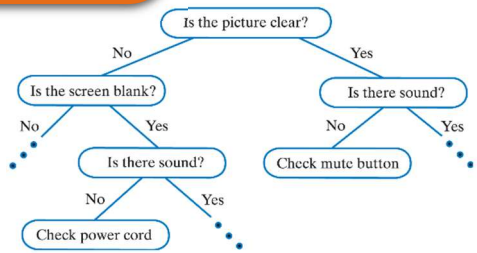
e.g. Dr. Peterson personality tests: Lots of similar questions in test, if participants answer all similarly = more likely they have that trait.

Decision Trees

Both Descriptive and Predictive.

Maps possible outcomes of series of related choices.

Useful for identifying criteria to divide data into classes, weighing possible actions against one another based on their potential outcomes.



Descriptive Learning Methods (Unsupervised)

1. **PCA – Principal Component Analysis**
    a. Help you identify which variables are important so you can compress the data by reducing the number of dimensions.
    b. *E.g. gender, days of week they typically shop, amount spent on average trip.*
2. **Cluster Analysis (K-means, moving centers)**
    a. Clusters similar data points to each other that are less like those in separate clusters
    b. *E.g. Consider a bookstore with different areas for the types of books: History, Self-Help, Romance, Mystery & Crime etc. The books in each of these clusters are more like each other than they are to other clusters.*
3. **Association Analysis (≡ Market Basket Analysis)**
    a. In a given set of records, each will contain a number of items. Association analysis allows you to determine the degree to which the items tend to be associated with one another.
4. **Neural Clustering (Kohonen Network (SOM))**
5. **Factor Analysis (CPA, CMA, CA)**
6. **Decision Trees (both Descriptive and Predictive).**

1. PCA – Principal Component Analysis

2. Cluster Analysis (K-means, Moving Centers)

Clustering is used for organizing variables into smaller, homogenous groups to simplify analysis and find patterns in data.

- Variables are grouped together based on similar traits, i.e. factors like age, income, education level etc.
- Typically a highly subjective process since it is dependent on one's familiarity with the data.

Cluster Types

Cluster Types	
Exclusive Clustering	Each variable can only belong in one single cluster. It cannot belong to any other cluster.
Fuzzy Clustering	Data points are assigned a probability belonging to one or more clusters
Overlapping Clustering	Each variable can belong to more than one cluster.
Hierarchical Clustering	Iterative process that either divides or combines clusters
Probabilistic Clustering	Data is clustered using algorithms that connect variables using distances or densities. This is done via computer.

K-Means

- Clustering algorithm
- **Useful for data that have NO labels, *i.e. data without defined categories or groups*.**
- Allows researcher to find groups of data which are like each other and cluster them together, even if the data is not labeled.
- Clusters defined by centroids, there are k centroids in each data set. k is a value returned by the algorithm (a point considered to belong to a particular cluster if the point is closer to that cluster's centroid than any other centroid).

- Iterative method, process of assigning data points to clusters based on current centroids and then choosing next, closer centroid is done until the data set reaches convergence.

Moving Centers

- Clustering algorithm to group data points
- Akin to K-Means method as they work in the same manner.

3. Association Analysis/Rules (≣ Market Basket Analysis)

- Best if used for finding the **most frequently occurring combinations** of variables in a data set.
- Commonly used for Market Basket Analysis
- *E.g. Amazon store has "Customers who viewed this item also viewed" and "frequently bought together".*
- Apriori is the main algorithm for detecting association rules.

4. Neural Clustering (Kohonen Network (Map))

- Classification method - AKA Self-Organizing Map (SOM)
- Unsupervised learning **neural network**
- In the end of the learning, all inputs are classified as the node they are closest to. Each input that is processed changes the location of the nodes.
- No variables to predict, it learns the structure of the data in order to distinguish clusters hidden within the data.
- Operates like a matrix that is made up of cells (nodes), vectors, and magnitudes (weights) of each vector.

Neural Network

Can be used for clustering as well as classifying data (predictive and descriptive, qualitative or quantitative dependent variable).

- Descriptive (Unsupervised) or Predictive (Unsupervised) learning
- The patterns recognized by neural networks are numerical and contained in vectors. Input data is translated from its raw form into numerical vector values.

+	Handles non-linear relations between the variables
-	**Requires massive amounts of computer power.**

5. Factor Analysis

- Collapses many variables into a few underlying factors.
- **Key concept: Multiple observed variables have similar patterns of responses because they are all associated with a latent *(i.e. not directly measured)* variable.**
- Involves grouping variables with similar attributes into a matrix using linear algebra techniques.
- In FA there are the same number of factors as there are variables.
- The eigenvalue is a measure of how much of variance of the observed variables a factor explains – any factor with an eigenvalue ≥ 1 explains more variance than a single observed variable.
- Factor analysis methods include: Principal Component Analysis (PCA), Correspondence Analysis (CA), and Multiple Correspondence Analysis (MCA).
    - **Correspondence Analysis (CA):** Intended for the analysis of qualitative variables.

        - Plots variables in relation to one another.
        - High frequency denotes a strong positive relationship, *i.e. two positively related categories, A&B, are opposed*, intermediate frequency denotes a weak relationship.
        - Can be used to convert qualitative values into quantitative values.
    - **Multiple Correspondence Analysis (MCA):** Component with > 2 variables.
        - Can be used to convert qualitative values to quantitative values.

6. Decision Trees (both Unsupervised/Descriptive & Supervised/Predictive

- Maps possible outcomes of series of related choices.
- Useful for identifying criteria to divide data into classes, weighing possible actions against one another based on their potential outcomes.
- Detects two-way interactions between tables.

+	**Easy to understand,** **Non-parametric,** **Models non-linear response of dependent variable,** **Can handle collinearity, missing data and outliers using isolating nodes,** **Simpler data preparation phase** **Reasonable computing times**
-	**Become extremely complex very quickly,** **Evaluates independent variables sequentially, not simultaneously, causing lack of robustness of the model and potential bias** (can be overcome by resampling and using a mean, but this will cause the loss of the simplicity of the model) **Large dataset required.**

Predictive Learning Methods (Supervised)

Multi-Layer Perceptron

A class of feed forward artificial network consisting of 3 layers of nodes: input layer, hidden layer, output layer. MLP uses a supervised learning technique called backpropagation for training.

Radial Basis Function

Neural network.

Works with only 1 hidden layer.

Uses circle as a basis for analysis

Prediction & classification.

Support Vector Machines

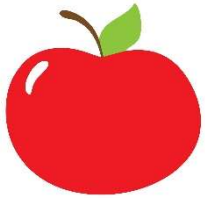
Used for classification, regression, and detecting outliers in data.

SVM is used to separate classes from each other.

Naïve Bayes Method

Classifies data based on the probability that it belongs to the class given certain features.

fruit has a high probability of being classified as an apple if it is red, round, and about 3 inches in diameter.



Linear Discriminant Analysis

Maximize separability: algorithm searches for a linear combination of variables (predictors) that best separates two classes (targets)

Similar to PCA, but instead of maximizing variation, maximizes separability.

Best when assumptions of homoscedasticity and multinormality are correct.

ANOVA

Regression Model

Decision Trees

Both Descriptive and Predictive.

Maps possible outcomes of series of related choices.

Useful for identifying criteria to divide data into classes, weighing possible actions against one another based on their potential outcomes.

K-Nearest Neighbors

Predictive Learning Methods (Supervised)

1. Multi-Layer Perceptron (MLP)

Neural Network Perceptron

The perceptron of an algorithm for supervised learning of binary classifiers.

- **Binary classifiers**: functions that can decide whether an input belongs to some specific class or not.
- Uses machine learning to maximize separability of variables into categories:
    a. Starts with random weights and sees how many it got right/wrong.
    b. Adjusts the weights and sees if it got more right this time than last time.
    c. Repeats until accurately separates the data into classes.
    d. Once model is trained, the weights are applied to a different dataset.

2. Radial Basis Function (RBF)

- Neural network.
- Works with only 1 hidden layer.
- Uses circle as a basis for analysis
- Prediction & classification.

-	**It may need large number of units in its hidden layer, which increases execution time of network without always yielding perfect modelling of complex structures and irregular data.**

3. Support Vector Machines

- Supervised learning methods used for classification, regression, and detecting outliers in data.
- Builds perceptive model that can assign new input objects into 1 category or another – Essentially SVM is used to separate classes from each other.

+	**Can model non-linear phenomena,** **Precision of prediction in certain cases,** **Robust**
-	**Does not directly provide estimated probabilities,** **Sensitive to choice of Kernel parameters,** **Long computation time** **Limited number of software programs that can implement it.**

4. Naive Beyes Method

- Classifies data based on the probability that it belongs to the class given certain features.
- **Can group objects based on features**: *e.g. fruit has a high probability of being classified as an apple if it is red, round, and about 3 inches in diameter.*

-	**Requires large data sets.**

5. Fisher's Linear Discriminant Analysis

- **Maximize separability**: algorithm searches for a linear combination of variables (predictors) that best separates two classes (targets)
- Similar to PCA, but instead of maximizing variation, maximizes separability.

- Best when assumptions of homoscedasticity and multinormality are correct.

+	**Fast calculation time,** **Concise models,** **Good at detecting global phenomena** (c.f. decision trees which detect local phenomena)**,** **Does not need a lot of records.**
-	**Only detects linear phenomena,** **Can only be used for continuous variables** (except for DISQUAL), **Sensitive to outliers.**

Disqual

- An extension of Fisher's Discriminant Analysis developed by Gilbert Saporta in 1975.
- Allows for qualitative dependent variables.

6. Regression Models

Linear regression

Assumes the following:

a. **Residuals** (error terms) are **normally distributed**.
b. **Best fitting regression line is a straight line**.
c. **Residuals** (error terms) have **constant variance** at every value of x.
d. **Residuals** (error terms) are **independent**.
e. **Residuals have a mean of zero** (error terms sum to 0).

7. Anova

8. K-nearest neighbors

9. Decision Trees are both predictive and descriptive

- Maps possible outcomes of series of related choices.
- Useful for identifying criteria to divide data into classes, weighing possible actions against one another based on their potential outcomes.
- Detects two-way interactions between tables.

+	**Easy to understand,** **Non-parametric,** **Models non-linear response of dependent variable,** **Can handle collinearity, missing data and outliers using isolating nodes,** **Simpler data preparation phase,** **Reasonable computing times.**
-	**Become extremely complex very quickly,** **Evaluates independent variables sequentially, not simultaneously, causing lack of robustness of the model and potential bias** (can be overcome by resampling and using a mean, but this will cause the loss of the simplicity of the model), **Large dataset required.**