Cheatography

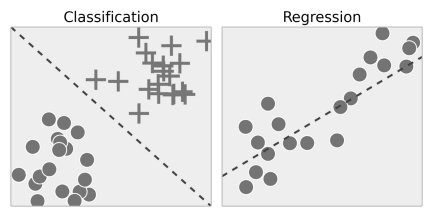
Machine Learning Model - Basics/Intermediate Cheat Sheet

by spriiprad via cheatography.com/122548/cs/22783/

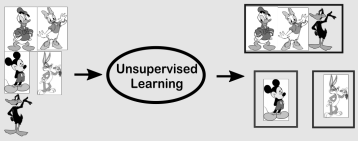
Supervised Vs Unsupervised Learnig

Supervised	Unsupervised
Used in Classification and Prediction	Dimension Reduction and clustering
Value of outcome must be known	No outcome variable to predict or classify
Learns from training data and applied to validation	No learning

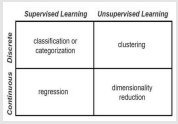
How Supervised Learning Looks



How Unsupervised Learning Looks



Supervised vs Unsupervised TLDR



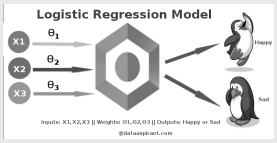
1. Linear Regression

Type of Response	Continuous
Simple Regression	**Multiple Regression**
One Independent Variable Used	Multiple Independent Variable Used
Only One Dependent Variable	Only One Dependent Variable

Relationships that are significant when using simple linear regression may no longer be when using multiple linear regression and vice-versa.

Insignificant relationships in simple linear regression may become significant in multiple linear regression.

2. How Logistic Regression Works



2. Logistic Regression

Type of Response	Categorical

It can be used for explanatory tasks (=profiling) or predictive tasks (=classification)

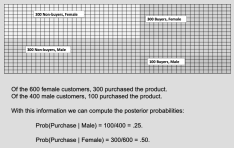
The predictors are related to the response Y via a nonlinear function called the logit

Reducing predictors can be done via variable selection

Types

1. Binary Regression	Two Categories.	Example: Spam or Not
2. Multinomial Logistic Regression	Three or more categories.	Example: Veg, Non-Veg, Vegan
3. Ordinal Logistic Regression	Three or more categories	Example: Movie rating from 1 to 5

3. How Naive Bayes Work



3. Naive Bayes Classifier

Type of Response	Categorical

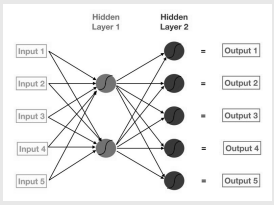
Probabilistic machine learning model that's used for classification task.

The heart of the classifier is based on the Bayes theorem.	Bayes theorem provides a way relating the likelihood of some outcome given some informative prior information.
We can find the probability of A happening, given that B has occurred.	B is the evidence and A is the hypothesis. That is presence of one particular feature does not affect the other.
Bayes Theorem Probability Formula	P(A/B) = (P(B|A)*P-(A))/P(B)
Naive Bayes works well when there is a large number of predictor variables	It also works when there are missing values.
The probability estimates are not very accurate	The classifications or predictions are generally accurate.

Assumptions

1. Predictors/features work independently on the target variable.	2. All the predictors have an equal effect on the outcome.

4. How Neural Net Works



4. Neural Networks

Type of Response	Both Categorical and Continuous (particularly useful)

Learns complex patterns using layers of neurons which mathematically transform the data.

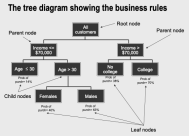
The layers between the input and output are referred to as “hidden layers”.

Learns relationships between the features that other algorithms cannot easily discover.

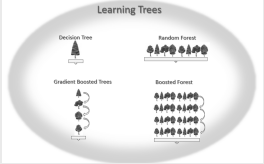
Architecture of Neural Net

Input Layer	Nodes(variables) with information from the external environment
Output Layer	Nodes(variables) that send information to the external environment or to another element in the network
Hidden Layer	Nodes that only communicate with other layers of the network and are not visible to the external environment

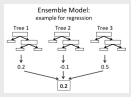
5. How Decision Trees Work



5. Different Types of Trees



5. How Ensemble Model Works

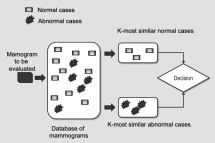


5. Decision Trees

The decision tree is produced by successively cutting the data set into smaller and smaller chunks, which are increasingly "pure" in terms of the value of the target variable.

Random Forest - Ensemble Method	**Boosted Trees - Ensemble Method**
Consists of a large number of individual decision trees that operate as an ensemble	Boosting is a method of converting weak learners into strong learners.
Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model’s prediction	Boosted trees is the process of building a large, additive tree by fitting a sequence of smaller trees
The predictions (and therefore the errors) made by the individual trees need to have low correlations with each other.	In boosting, each new tree is a fit on a modified version of the original data set.
Random Forests train each tree independently, using a random sample of the data.	GBTs train one tree at a time, where each new tree helps to correct errors made by previously trained trees.

6. How KNN works



6. K-Nearest Neighbors

Type of Response	Both Categorical and Continuous

KNN is method for classifying objects based on their similarity to a data with known classifications.

K-Nearest Neighbors (KNN) makes a prediction for a new observation by searching for the most similar training observations and pooling (usually done by taking the mean average) their values

Training set has to be very large for this to work effectively

Redundant and/or irrelevant variables can distort the classification results; the method is sensitive to noise in the data.

Nominal variables pose problems for measuring distance

It is a non-parametric model ... does not require distribution assumptions regarding the variables and does not make statistical inferences to a population

KNN is an example of a family of algorithms known as instance-based or memory-based learning that classify new objects by their similarity to previously known objects.