Introduction to Machine Learning

With Scikit-Learn and OpenCV

Machine Learning + Big Data in Real Time + Cloud Technologies

=> The Future of Intelligent Systems

Where to Find The Code and Materials?

https://github.com/iproduct/course-ml

Agenda for This Lesson - I

- Machine learning
- Supervised, Semi-supervised and unsupervised ML
- Classification, clustering and reinforcement learning
- Mean, median, standard deviation, percentiles, histograms.
- Normal distribution
- Data visualization using Matplotlib
- Training-testing-evaluation

Agenda for This Lesson - II

Practical ML using Scikit-learn and Python:

- Prediction: linear regression, polynomial regression, multiple regression
- Classification: logistic regression, K Nearest Neighbors (KNN)
- Gradient descent
- Clustering : K-Means Clustering
- Principal Component Analysis (PCA) & Feature Selection
- Linear discriminant analysis – LDA
- Support Vector Machines (SVM). Kernels
- Decision Trees. Random Forests

Introduction to Machine Learning

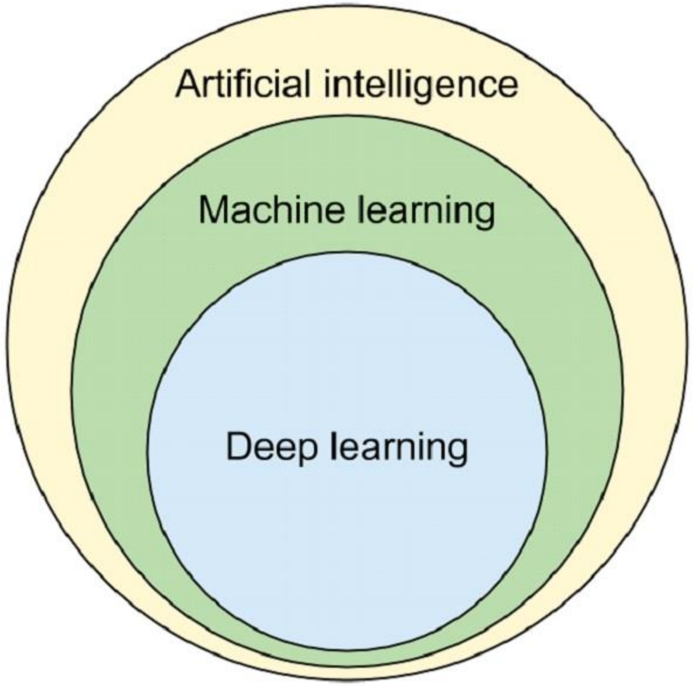
What Is Machine Learning?

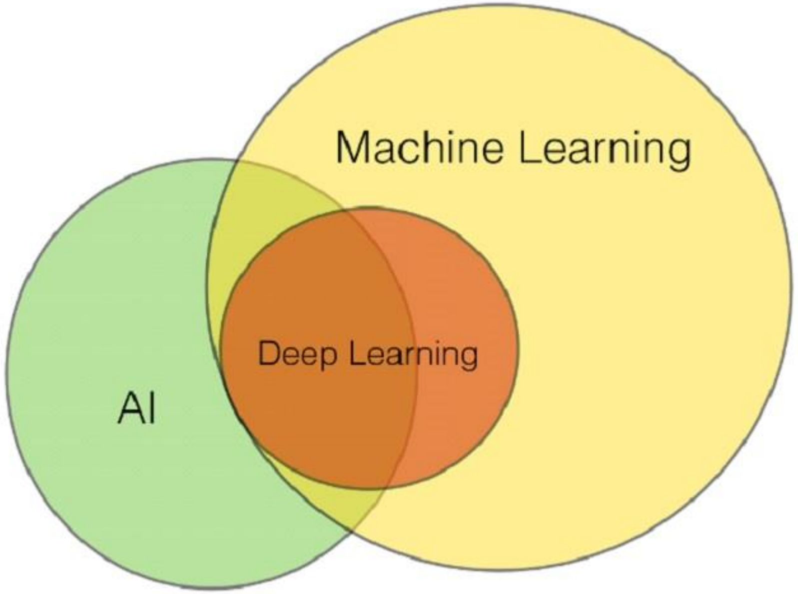
- **Machine Learning (ML)** –the study of computer algorithms that improve automatically through experience. A subset of Artificial Intelligence (AI).
- ML algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so.
- ML algorithms are used in a wide variety of applications, where it is difficult or unfeasible to develop conventional algorithms for the task – e.g. computer vision, email filtering, digital assistants, natural language processing (NLP), recommendations, personalized online advertising, chatbots, fraud detection, cybersecurity, medical image analysis, self-driving cars, self-driving databases, and much, much more ...
- ML is closely connected with computational statistics, mathematical optimization, data mining, predictive analytics.

Machine Learning and Artificial Intelligence

- ML learns and predicts based on passive observations, whereas AI implies an agent interacting with the environment to learn and take actions that maximize its chance of successfully achieving its goals.

Judea Pearl, The Book of Why, 2018





Types of Learning Algorithms: Supervised Learning

- **Supervised learning** – build a mathematical model of a set of data that contains both the inputs and the desired outputs.
- Training data – a set (matrix) of training examples (feature vectors), having one or more inputs and the desired output. Through iterative optimization of an objective function, supervised learning algorithms learn a function that can be used to predict the output associated with new inputs.
- Examples: active learning, classification and regression.
  - Active learning – learning algorithm can interactively query a user (teacher or oracle) to label new data points with the desired outputs (optimal experimental design).
  - Classification algorithms – when the outputs are restricted to a limited set of values
  - Regression algorithms – when the outputs may have any numerical value within a range.
  - Similarity learning – using a similarity function measuring how similar two objects are – ranking, recommender systems, visual identity tracking, face and speaker verification.

Types of Learning Algorithms: Unsupervised Learning

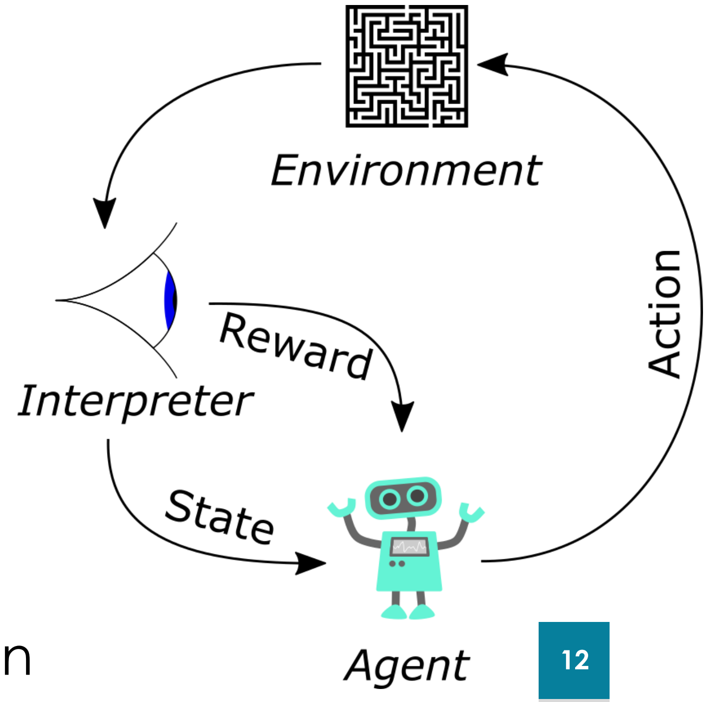
- **Unsupervised learning** – takes a data set that contains only inputs, and find structure in the data, like grouping or clustering of data points by identifying data commonalities
- Algorithms learn from test data that has not been labeled, classified or categorized.
- Applications: density function estimation, summarizing and explaining data features.
- Cluster analysis is the assignment of a set of observations into subsets (called clusters) so that observations within the same cluster are similar according to one or more predesignated criteria, while observations drawn from different clusters are dissimilar.
- Different clustering techniques make different assumptions on the structure of the data, often defined by some similarity metric and evaluated, for example, by internal compactness, or the similarity between members of the same cluster, and separation, the difference between clusters. Other methods are based on estimated density and graph connectivity.

Types of Learning Algorithms: Semi-supervised Learning

- **Semi-supervised learning** – falls between unsupervised learning (without any labeled training data) and supervised learning (with completely labeled training data). During training, it uses a smaller labeled data set to guide classification and feature extraction from a larger, unlabeled data set.
- In **weakly supervised learning**, the training labels are noisy, limited, or imprecise; however, these labels are often cheaper to obtain, resulting in larger effective training sets.

Types of Learning Algorithms: Reinforcement Learning

- **Reinforcement learning** – concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward
- Due to its generality, the field is studied in many other disciplines, such as game theory, control theory, operations research, information theory, simulation-based optimization, multi-agent systems, swarm intelligence, statistics and genetic algorithms.
- In ML, the environment is typically represented as a Markov decision process (MDP). Many reinforcement learning algorithms use dynamic programming techniques.
- Reinforcement learning algorithms do not assume knowledge of an exact mathematical model of the MDP, and are used when exact models are infeasible.
- Examples: reinforcement learning algorithms are used in autonomous vehicles or in learning to play a game against human



Types of Learning Algorithms: Self Learning

- **Self learning** – as a machine learning paradigm was introduced in 1982 along with a neural network capable of self-learning named crossbar adaptive array (CAA). It is a learning with no external rewards and no external teacher advice.
- The CAA self-learning algorithm computes, in a crossbar fashion, both decisions about actions and emotions (feelings) about consequence situations. The system is driven by the interaction between cognition and emotion.
- In situation s perform an action a; Receive consequence situation s'; Compute emotion of being in consequence situation $v(s')$; Update crossbar memory $w'(a,s) = w(a,s) + v(s')$.
- The backpropagated value (secondary reinforcement) is the emotion toward the consequence situation. The CAA exists in two environments, one is the behavioral environment where it behaves, and the other is the genetic environment, wherefrom it initially and only once receives initial emotions about situations to be encountered in the behavioral environment.

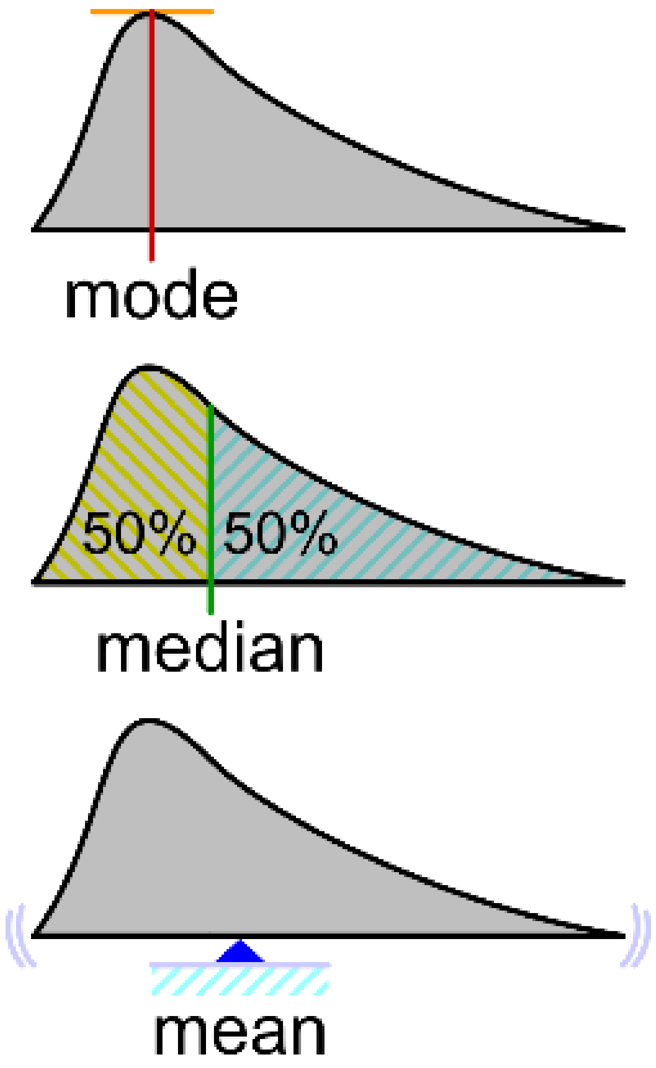
Types of Learning Algorithms: Feature learning – I

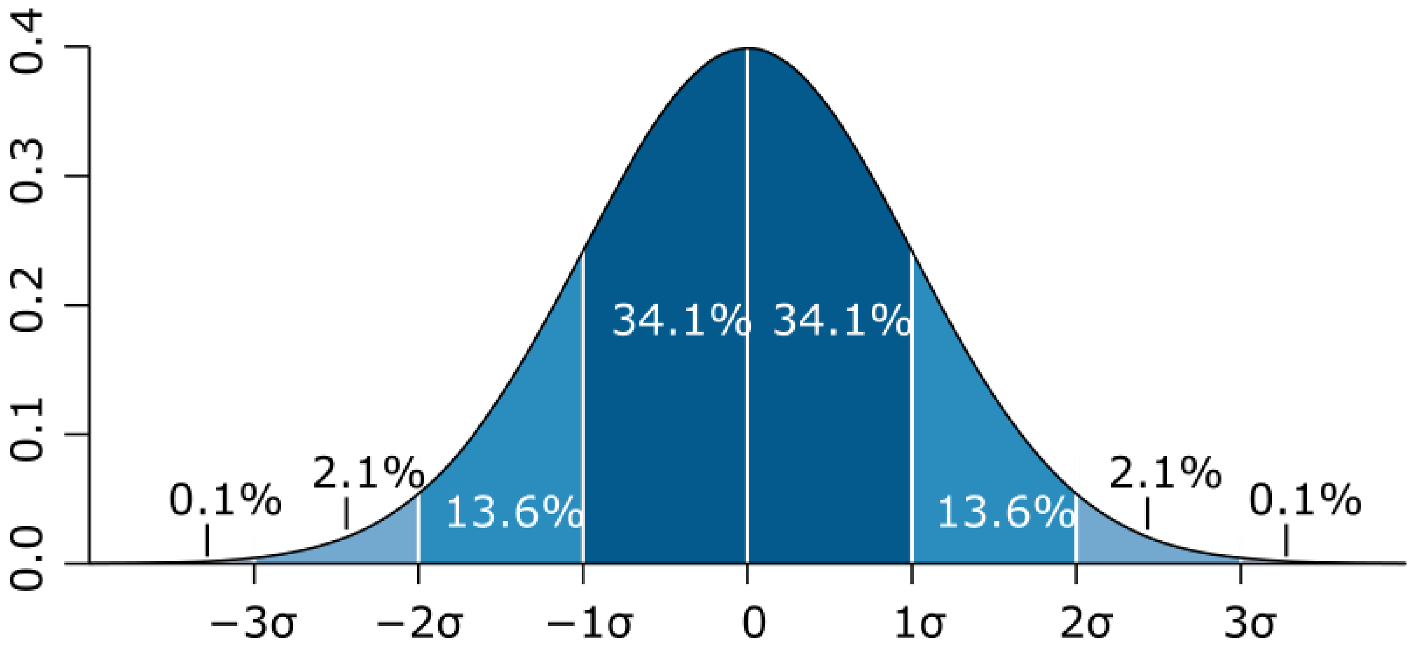
- **Feature learning (representation learning algorithms)** – discover better representations of the inputs provided during training. Examples: principal components analysis and cluster analysis.
- Attempt to preserve the information in their input but also transform it in a way that makes it useful, often as a pre-processing step before performing classification or predictions.
- Allows a machine to both learn the features and use them to perform a specific task.
- Feature learning can be either supervised or unsupervised. In supervised feature learning, features are learned using labeled input data. Examples: artificial neural networks, multilayer perceptrons, and supervised dictionary learning. In unsupervised feature learning, features are learned with unlabeled input data. Examples: dictionary learning, independent component analysis, autoencoders, matrix factorization, and clustering.

Types of Learning Algorithms: Feature learning – II

- Manifold learning algorithms attempt to do so under the constraint that the learned representation is low-dimensional.
- Sparse coding algorithms attempt to do so under the constraint that the learned representation is sparse, meaning that the mathematical model has many zeros.
- Multilinear subspace learning algorithms aim to learn low-dimensional representations directly from tensor representations for multidimensional data, without reshaping them into higher-dimensional vectors.
- Deep learning algorithms discover multiple levels of representation, or a hierarchy of features, with higher-level, more abstract features defined in terms of (or generating) lower-level features. It has been argued that an intelligent machine is one that learns a representation that disentangles the underlying factors of variation that explain the observed data.

Mean, Median, Mode, Standard Deviation

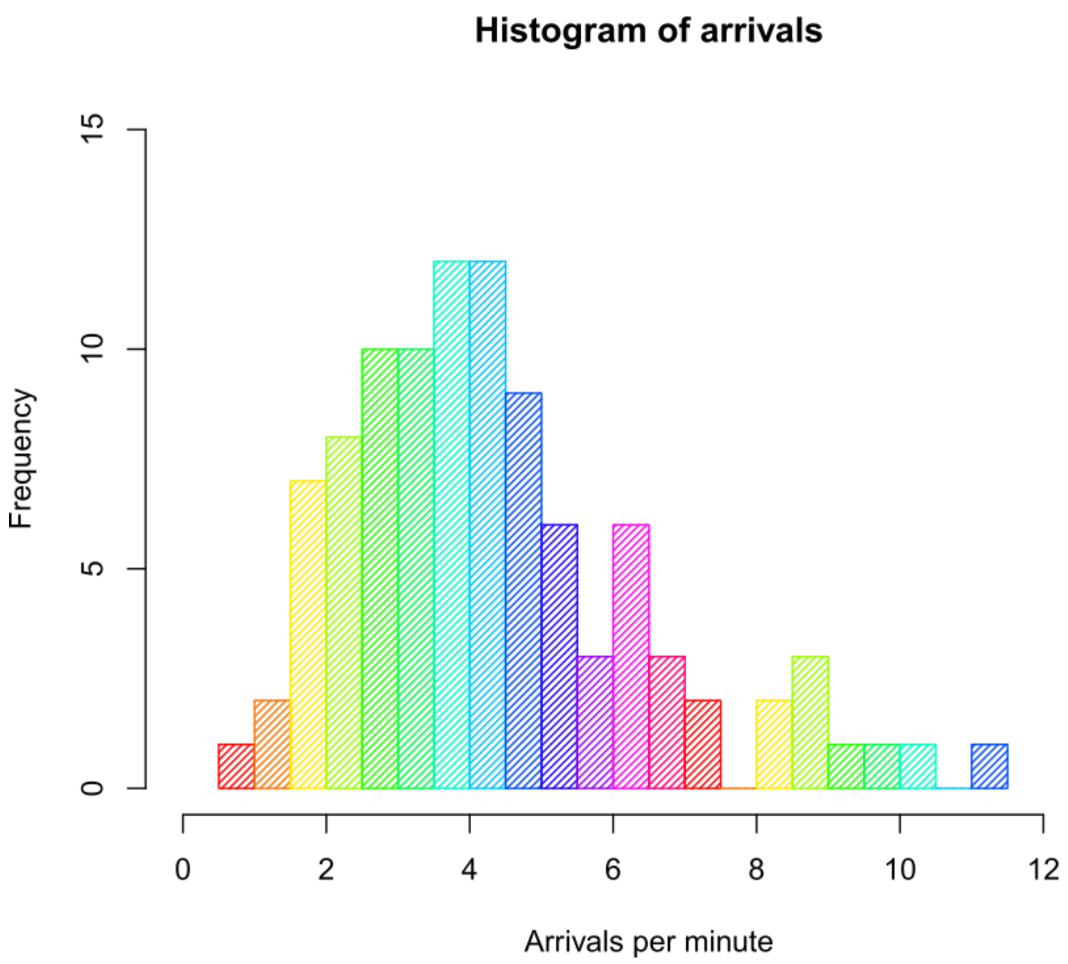




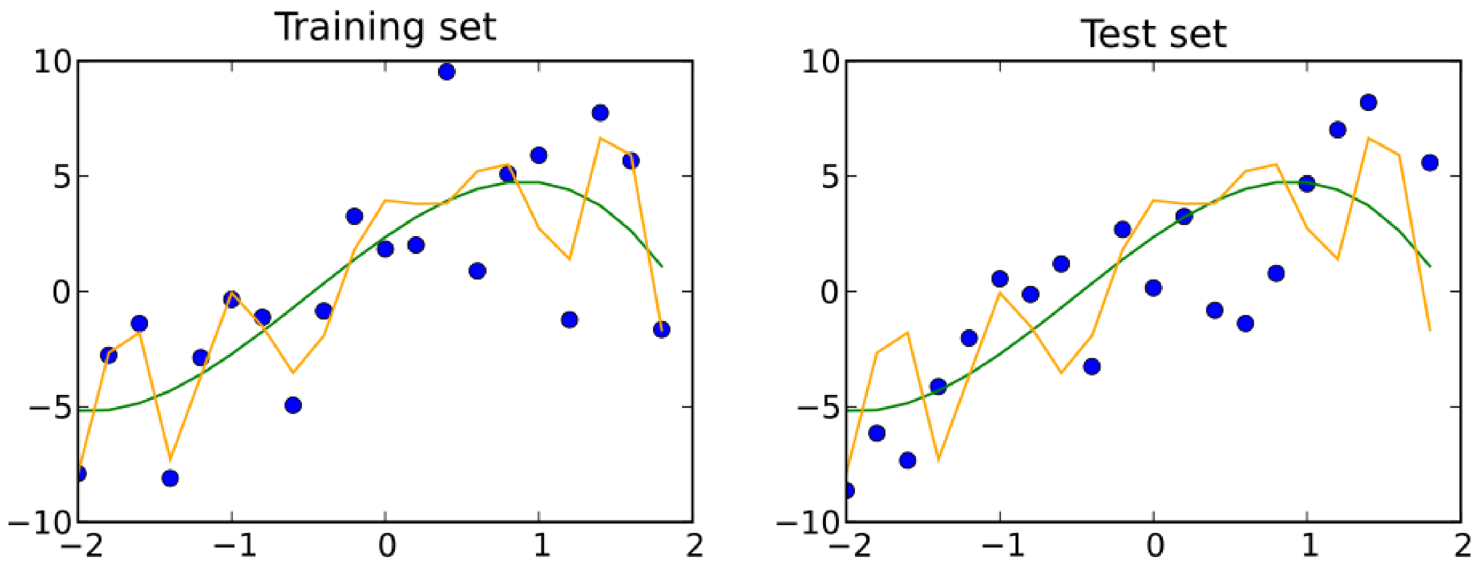
$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

Percentiles. Histograms

- Percentile (or a centile) is a type of quantile which divides the given probability distribution, or sample, into 100 equal-sized intervals; this allows the data to be analyzed in terms of percentages. For example, the 20th percentile is the value (or score) below which 20% of the observations are found, and above which 80% are found.
- Histogram – approximate representation of the distribution of numerical data. To construct a histogram, the first step is to "bin" (or "bucket") the range of values—that is, divide the entire range of values into a series of intervals—and then count how many values fall into each interval.



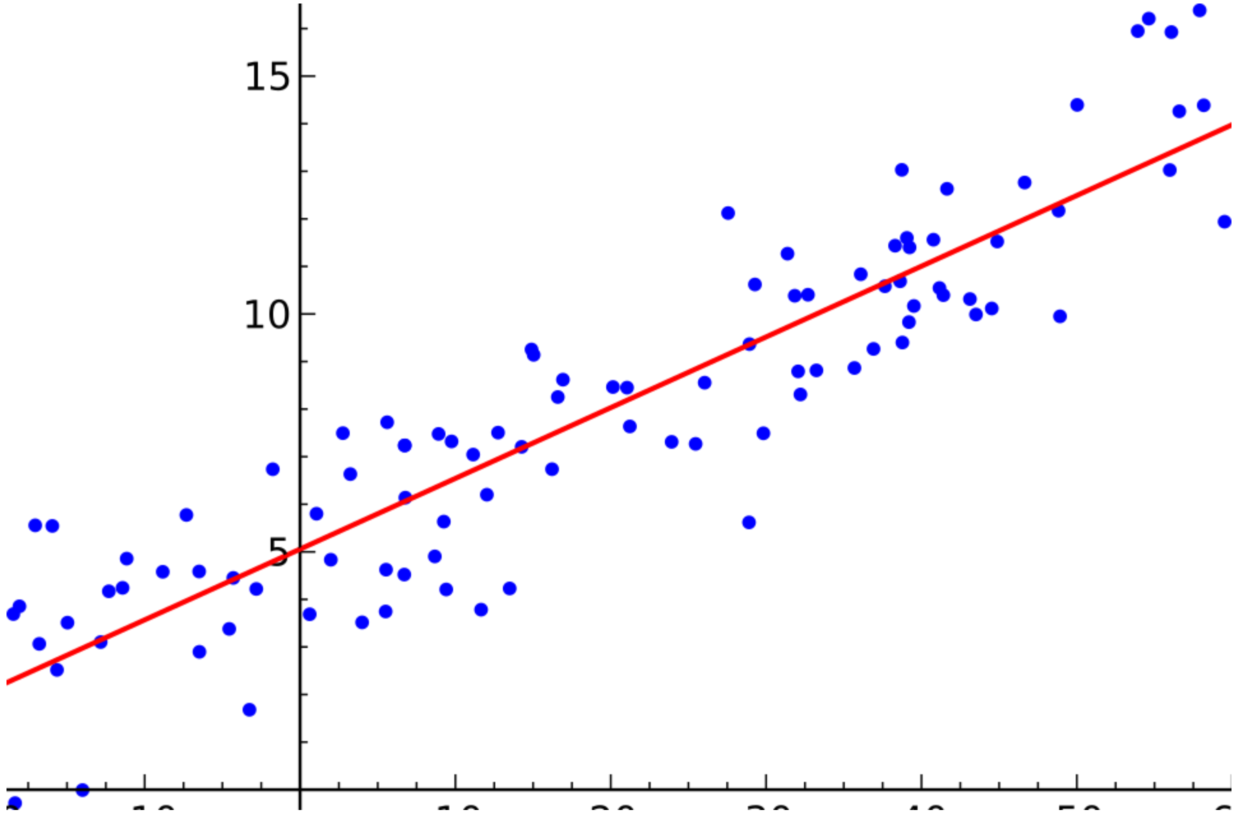
Training and Testing Data Sets

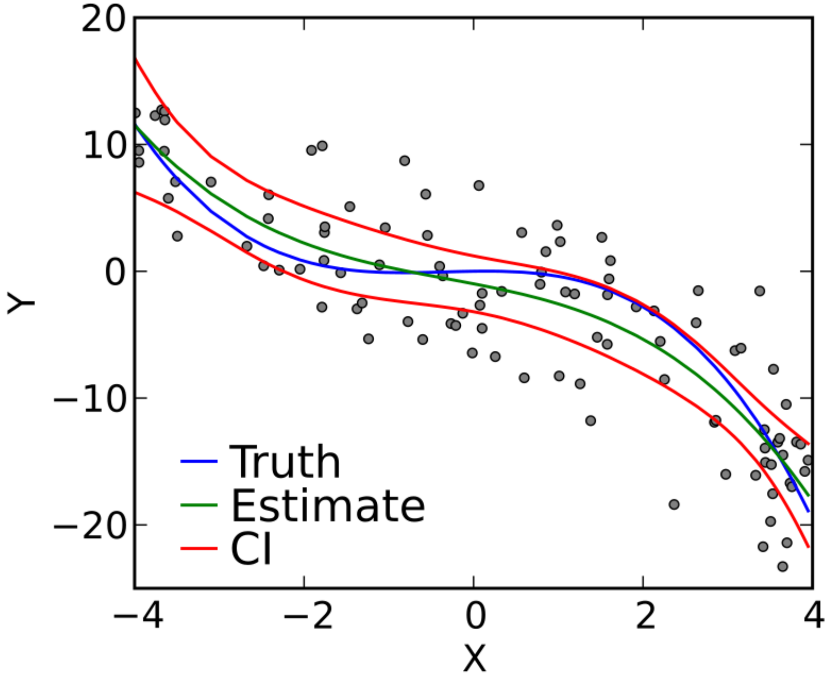


$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

A training set (left) and a test set (right) from the same statistical population are shown as blue points. Two predictive models are fit to the training data. Both fitted models are plotted with both the training and test sets. In the training set, the MSE of the fit shown in orange is 4 whereas the MSE for the fit shown in green is 9. In the test set, the MSE for the fit shown in orange is 15 and the MSE for the fit shown in green is 13. The orange curve severely overfits the training data, since its MSE increases by almost a factor of four when comparing the test set to the training set. The green curve overfits the training data much less, as its MSE increases by less than a factor of 2.

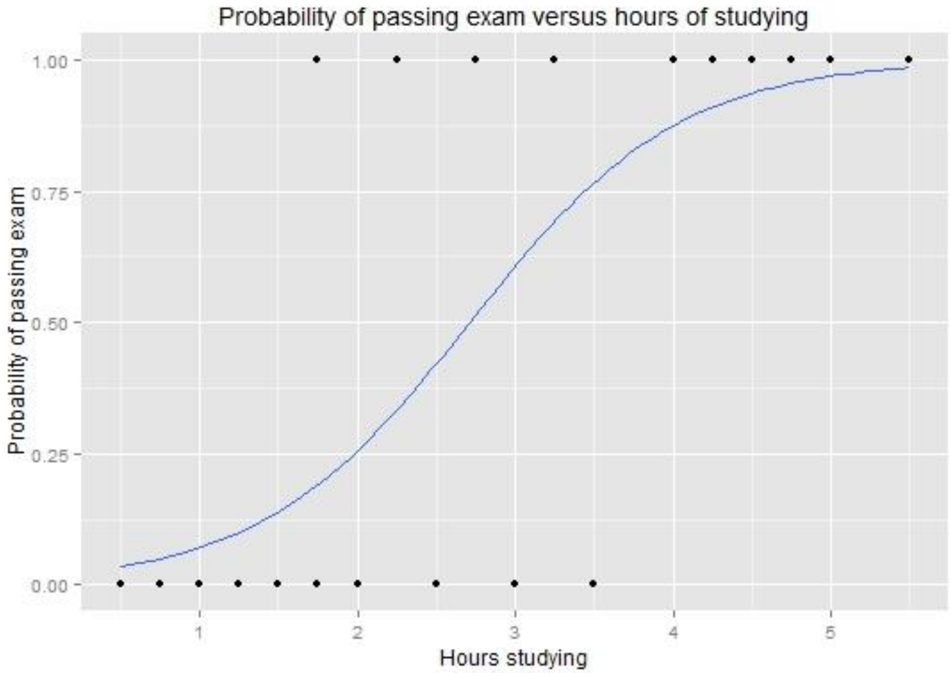
Prediction: Linear Regression, Polynomial Rregression

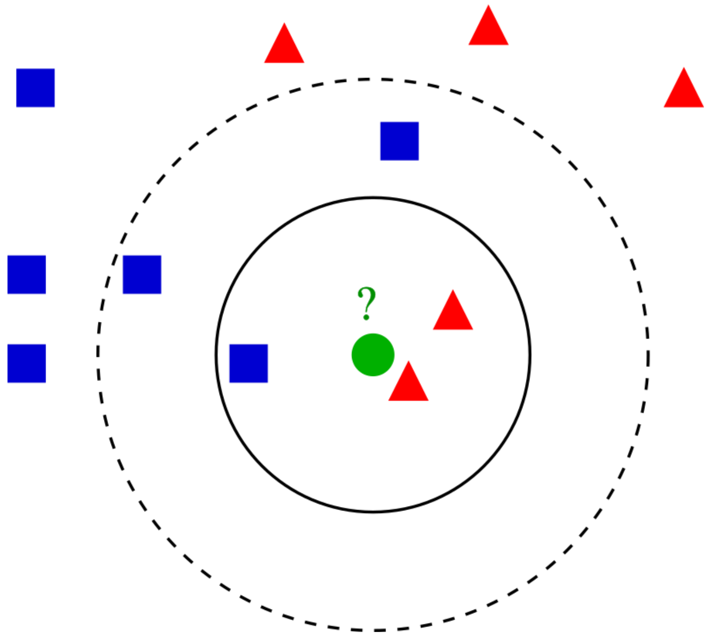




$$Y_i = f(X_i, \beta) + e_i$$

Classification: logistic regression, K Nearest Neighbors





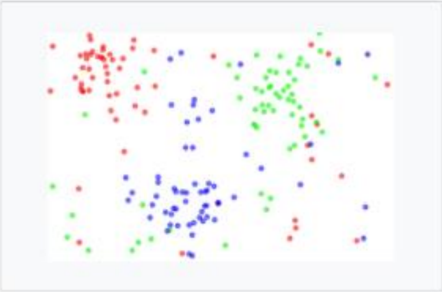


Fig. 1. The dataset.

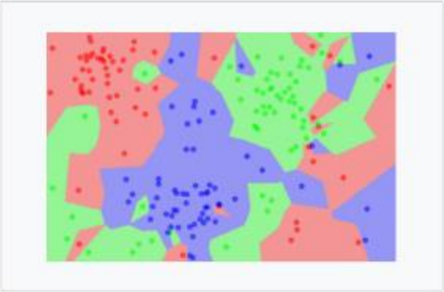


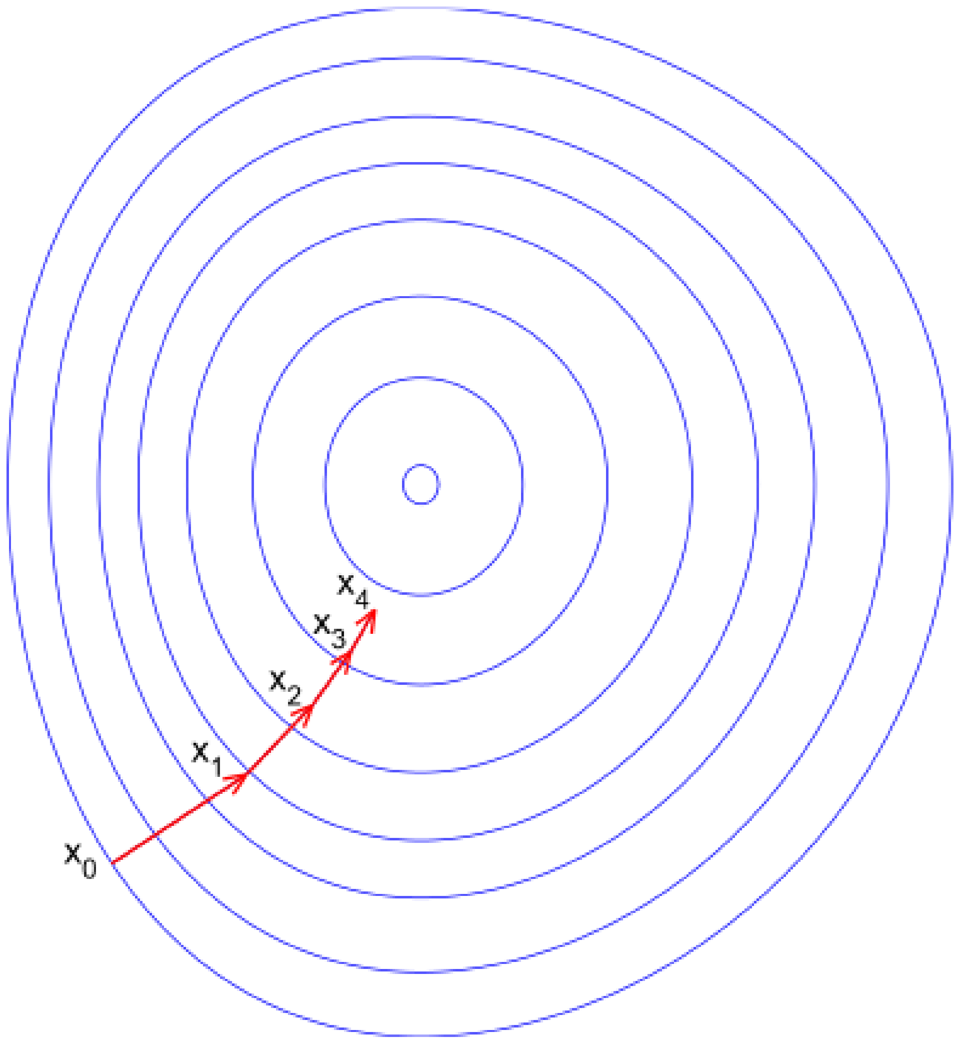
Fig. 2. The 1NN classification map.

Fig. 3. The 5NN classification map.

Fig. 4. The CNN reduced dataset.

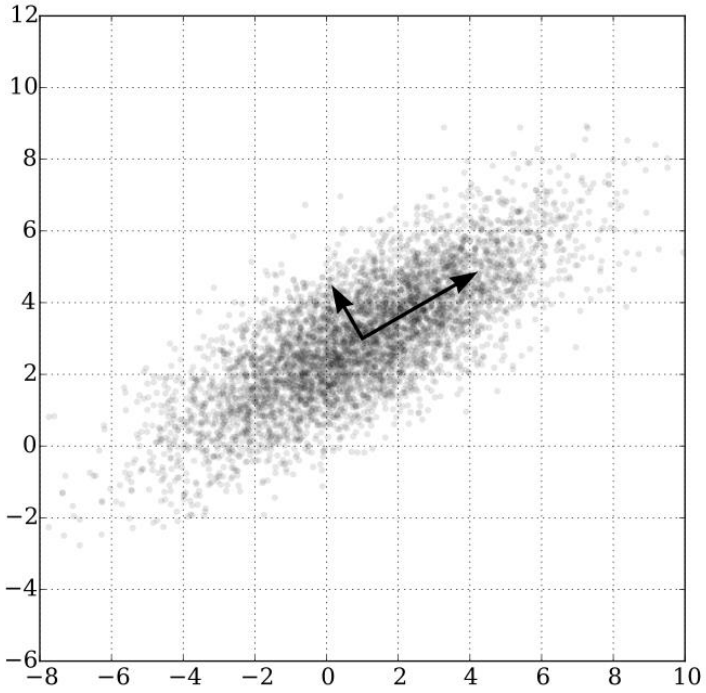
Fig. 5. The 1NN classification map based on the CNN extracted prototypes.

Gradient Descent



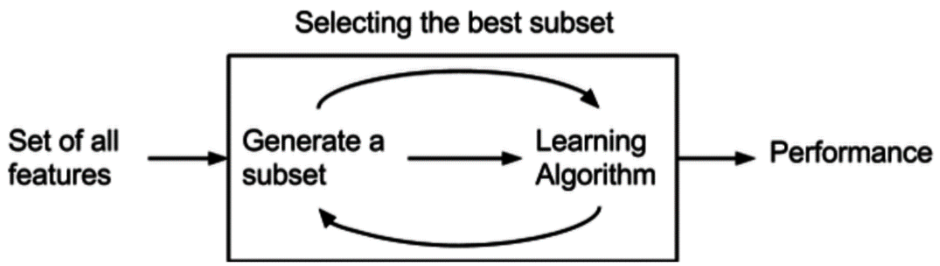
Principal Component Analysis (PCA)

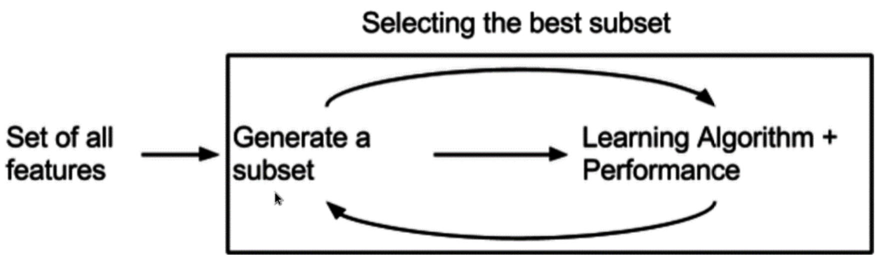
- Principal components of a collection of points in a real p-space are a sequence of direction vectors, where the vector is the direction of a line that best fits the data while being orthogonal to the first vectors. Here, a best-fitting line is defined as one that minimizes the average squared distance from the points to the line. These directions constitute an orthonormal basis in which different individual dimensions of the data are linearly uncorrelated.
- Principal component analysis (PCA) is the process of computing the principal components and using them to perform a change of basis on the data, sometimes using only the first few principal components and ignoring the rest.



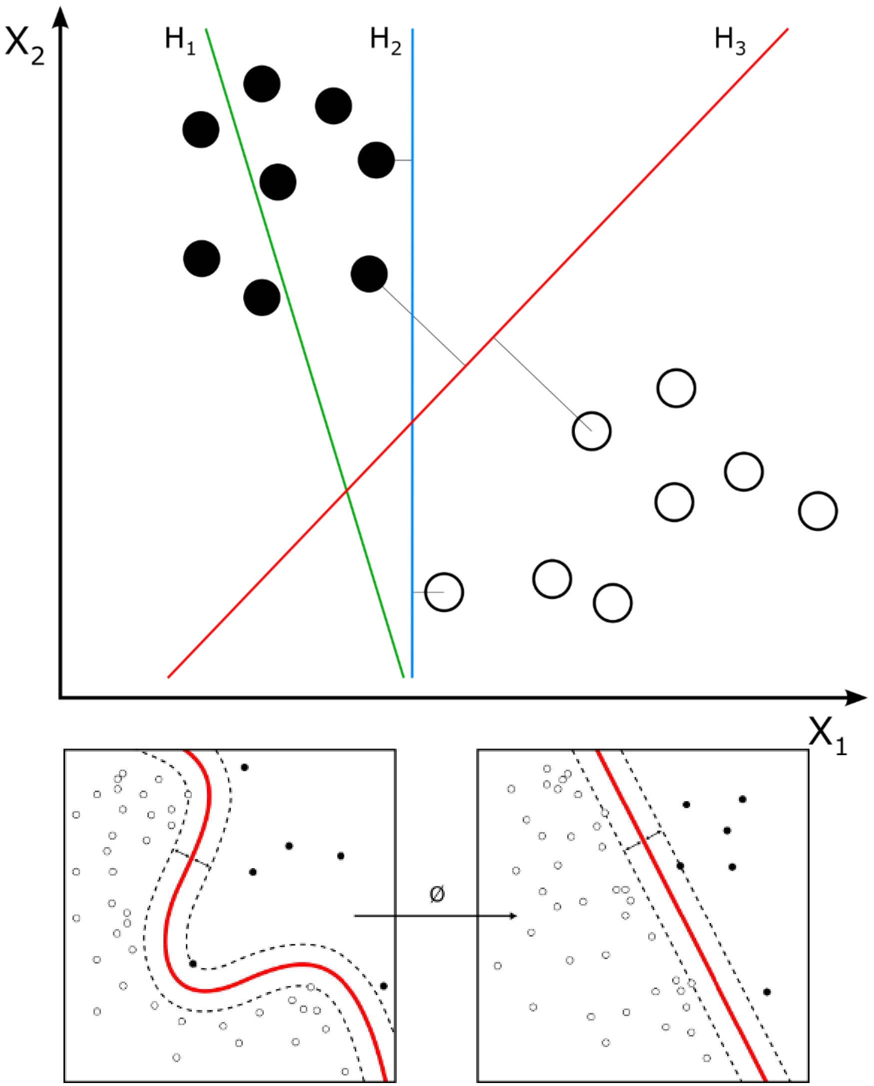
Feature Selection

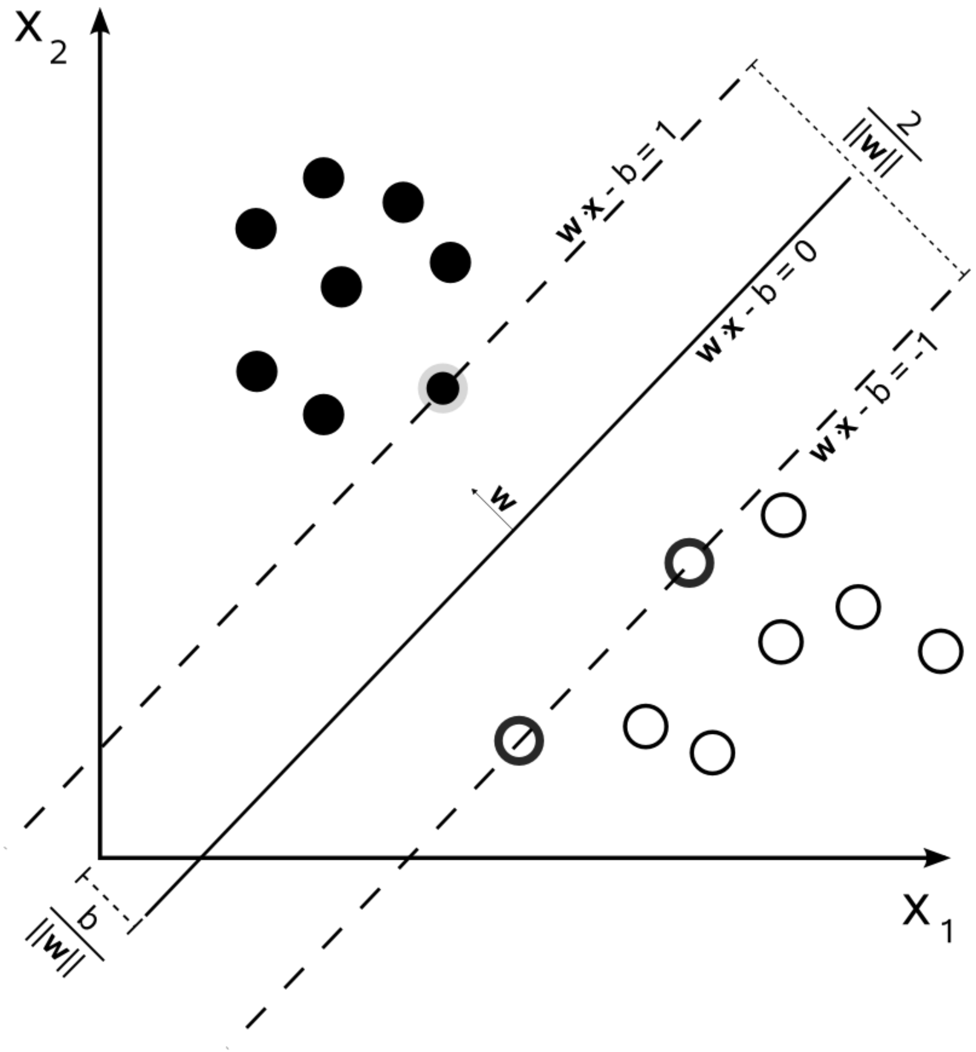
- Feature selection (variable selection, attribute selection) - a process of selecting a subset of relevant features (variables, predictors) for use in model construction by removing features that are redundant or irrelevant. Reasons to use:
    - simplification of models to make them easier to interpret by researchers/users;
    - shorter training times;
    - to avoid the curse of dimensionality;
    - enhanced generalization by reducing (formally, reduction of variance)
- Overview on metaheuristics methods – filter, wrapper and embedded methods:



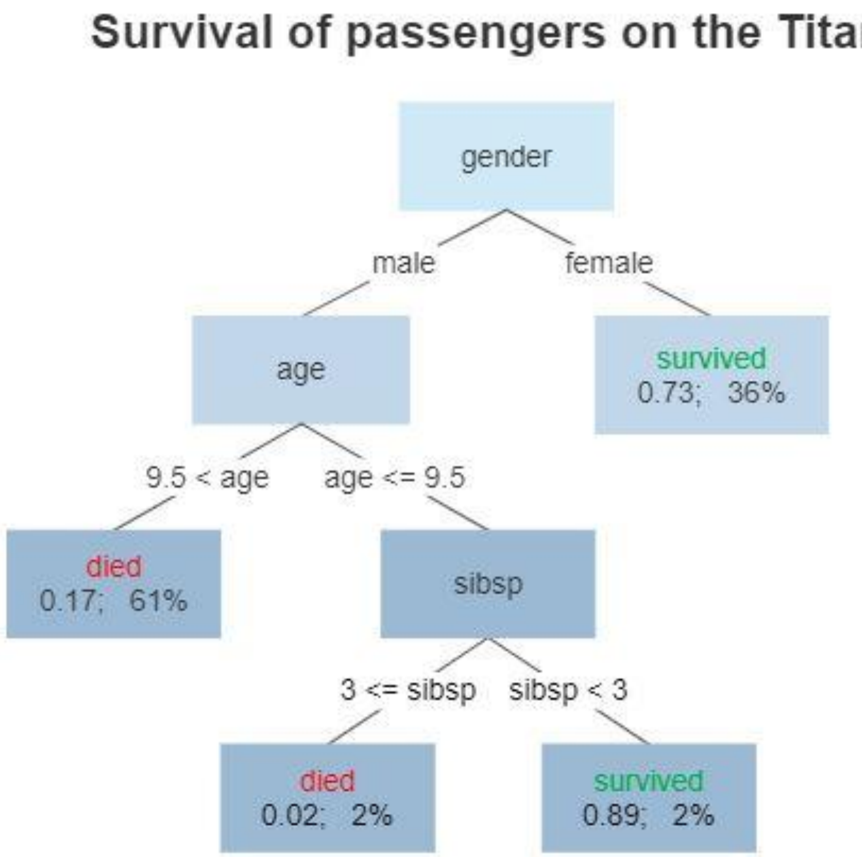


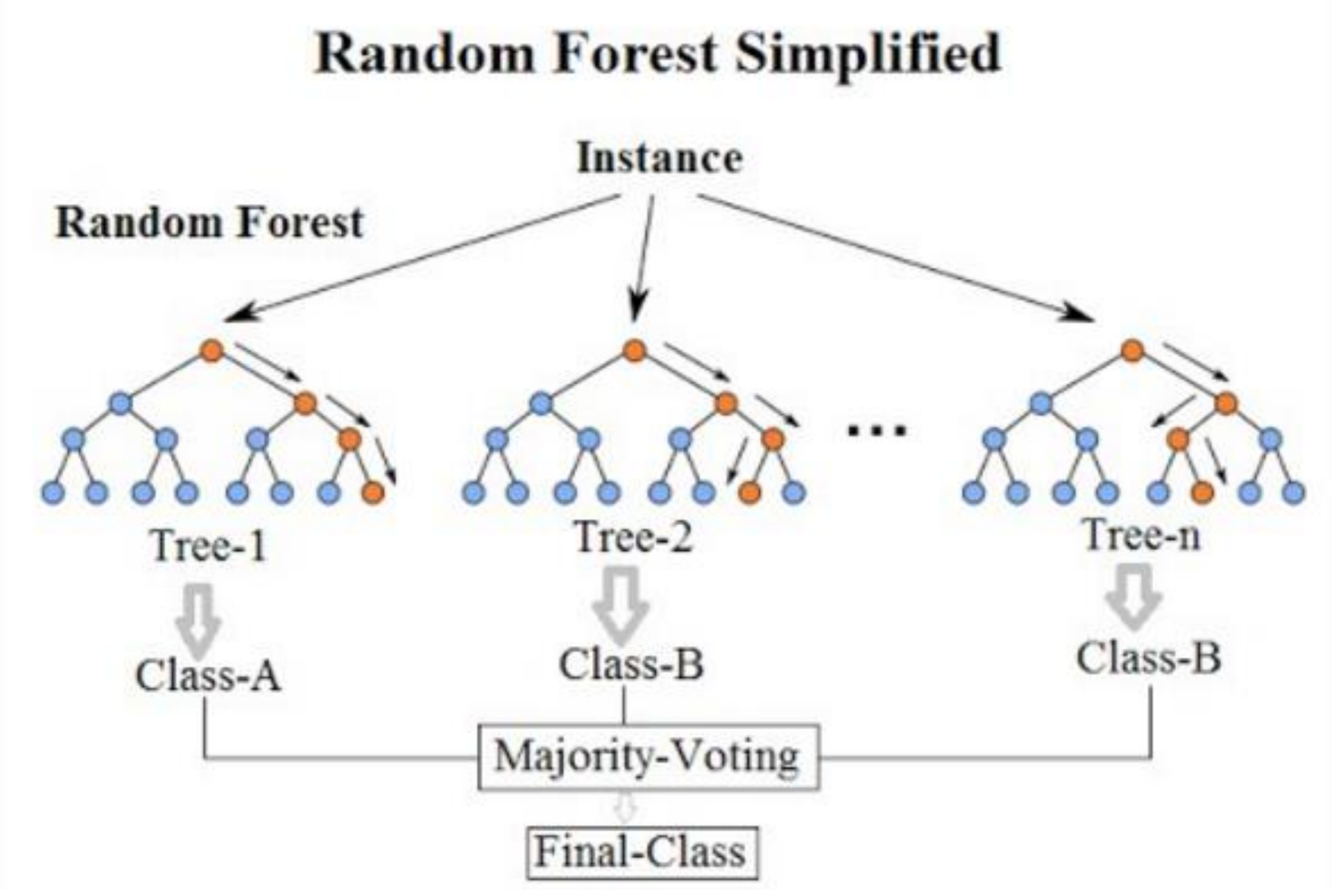
Support Vector Machines (SVM).





Decision Trees. Random Forests





Thank's for Your Attention!

Trayan Iliev

IPT – Intellectual Products & Technologies

http://iproduct.org/

https://github.com/iproduct

https://twitter.com/trayaniliev

https://www.facebook.com/IPT.EACAD