Session -I-
Machine Learning

asst.prof.dr. Boran Şekeroğlu

18/04/2019

Asst.Prof.Dr. Boran Şekeroğlu

Outline

- Learning
- Introduction
- Learning Types
- Classification vs. Regression
- Feature Engineering
- Model Performance Assesment
- Machine Learning Techniques

Asst.Prof.Dr. Boran Şekeroğlu

WHAT IS LEARNING?

- understanding and making connections between prior and new knowledge, independent and critical thinking and ability to transfer knowledge to new and different contexts.
  - Queens University Scienctists

Asst.Prof.Dr. Boran Şekeroğlu

Can Machines or Computers Learn?

- Computers will be able to accurately predict the outcome of the Israeli-Palestinian conflict. Feed the computer enough data on a number of "parallel universes," and the computer will be capable of observing the implications of each of these universes and then find patterns, allowing predictions to be made about the future of the conflict.
  - Dr. Kira Radinsky, (a computer scientist and "machine learning" expert), argued in the Israeli newspaper "Ha'aretz" (2015)

Can Machines or Computers Learn?

- Computers are not “creative,” do not “learn” and cannot “predict.” Computers can only be tasked with making inductive predictions based on past experiences. They can then seek complex correlations in the databases in order to present them as “Actionable Insights.”
  - Joab Rosenberg (2016)

Machine Learning – Introduction

- Machine learning is a sub-field of computer science that is concerned with building algorithms which, to be useful, rely on a collection of examples of some phenomenon.

- These examples can come from nature, be handcrafted by humans or generated by another algorithm.

Asst.Prof.Dr. Boran Şekeroğlu

Machine Learning – Introduction

- Machine learning can also be defined as the process of solving a practical problem by:
  - gathering a dataset, and
  - algorithmically building a statistical model based on that dataset.

What are the Types of Learnings in Machine Learning?

- 1. Supervised Learning:
  - In supervised learning, the dataset is the collection of **labeled** examples $\{(x_i, y_i)\}_{i=1}^{N}$.
  - Each element x_i among N is called a feature vector.

Asst.Prof.Dr. Boran Şekeroğlu

Supervised Learning

- The goal of a supervised learning algorithm is to use the dataset to produce a model that takes a feature vector x as input and outputs information that allows deducing the label for this feature vector.
- Example:
  - Input images and their corresponding labels:

Asst.Prof.Dr. Boran Şekeroğlu

Unsupervised Learning

- In unsupervised learning, the dataset is a collection of **unlabeled** examples $\{(x_i)\}_{i=1}^{N}$.

- Again, x is a feature vector.

- The goal of an unsupervised learning algorithm is to create a model that takes a feature vector x as input and either transforms it into another vector or into a value that can be used to solve a practical problem.

Unsupervised Learning

- Example:
  - Input images without labels:

Semi-Supervised Learning

- In semi-supervised learning, the dataset contains both labeled and unlabeled examples.
- Usually, the quantity of unlabeled examples is much higher than the number of labeled examples.
- The goal of a semi-supervised learning algorithm is the same as the goal of the supervised learning algorithm

Reinforcement Learning

- Reinforcement learning is a sub-field of machine learning where the machine "lives" in an environment and is capable of perceiving the state of that environment as a vector of features.
- The machine can execute actions in every state. Different actions bring different rewards and could also move the machine to another state of the environment.
- The goal is to learn a **policy**. A policy is a function *f* (similar to the model in supervised learning) that takes the feature vector of a state as input and outputs an optimal action to execute in that state.

Classification vs. Regression: Classification

- Classification is a problem of automatically assigning a label to an unlabeled example.
- In machine learning, the classification problem is solved by a classification learning algorithm that takes a collection of labeled examples as inputs.
- In a classification problem, a label is a member of a **finite set** of classes.

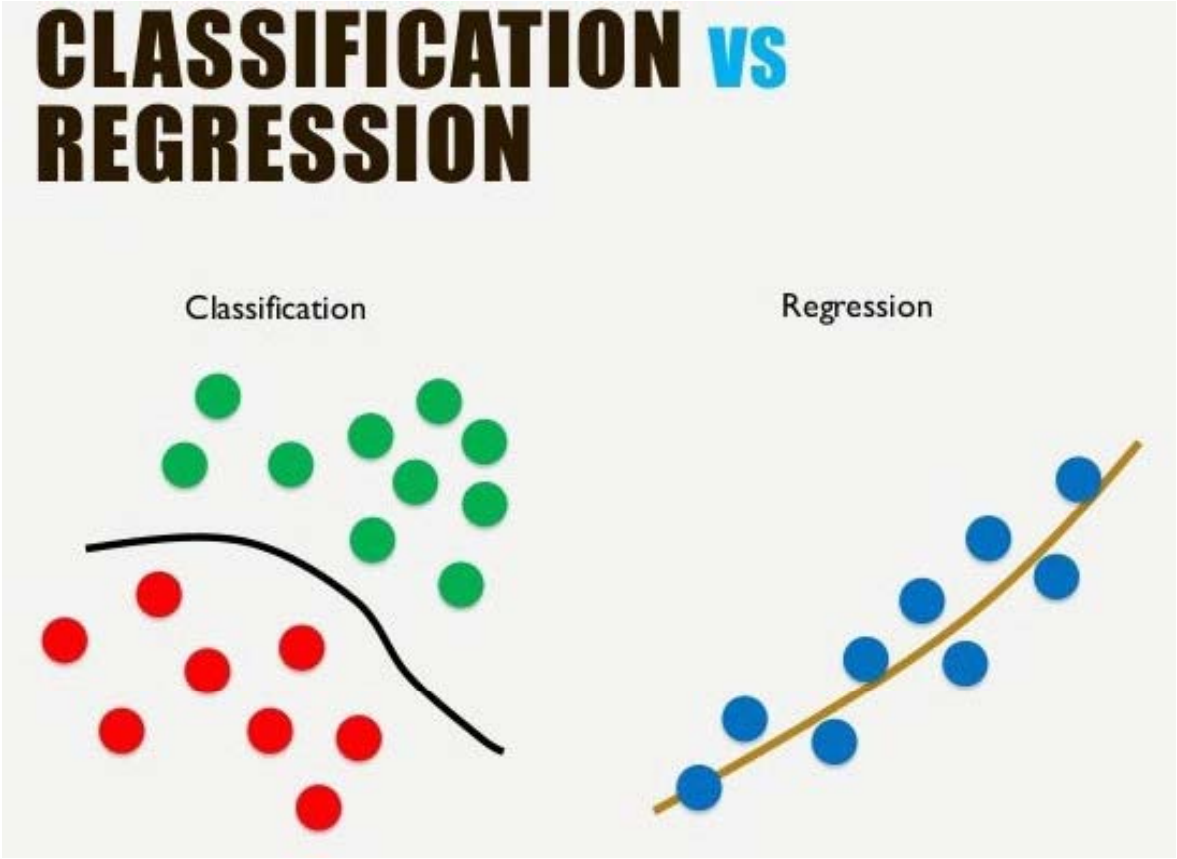
Classification vs. Regression: Classification

- If the size of the set of classes is two ("sick"/"healthy", "benign"/"malignant"), we talk about **binary** classification (also called **binomial** in some books.
- **Multiclass** classification (also called **multinomial**) is a classification problem with **three or more** classes. (Ex: dataset contains 16 classes with normal case, right bundle brunch block, atrial fibrillation etc.)

Classification vs. Regression: Regression

- Regression is a problem of predicting a real-valued label (often called a target) given an unlabeled example.
- The regression problem is solved by a regression learning algorithm that takes a collection of labeled examples as inputs and produces a model that can take an unlabeled example as input and output a target

Classification vs. Regression: Regression



Asst.Prof.Dr. Boran Şekeroğlu

Feature Engineering

- For any Machine Learning Application or Research, you need a “dataset”.

- Dataset can be from “images”, “numerical data” or “categorical data”.

  - Images can be benign or malignant images of skin cancer.
  - Numerical and Categorical data can be blood test results of patients.

Feature Engineering

- The problem of transforming raw data into a dataset is called **feature engineering.**
- **Everything measurable can be used as a feature.**

Feature Engineering – one-hot encoding

- When some feature in your dataset is categorical, like “Moderate”, “normal”, “Elevated,” or “Poor”, you can transform such a categorical feature into several binary ones.

Moderate:	[1 0 0 0]
Normal:	[0 1 0 0]
Elevated:	[0 0 1 0]
Very Poor:	[0 0 0 1]

Feature Engineering – Binning

- An opposite situation, occurring less frequently in practice…
  - Blood Test Results can be binning (bucketing) into a bin:
    - 0 – 50 : [1 0 0]
    - 50 – 100 : [0 1 0]
    - >100: [0 0 1]

Feature Engineering – Data Normalization

- Normalization is the process of converting an actual range of values which a numerical feature can take, into a standard range of values, typically in the interval [-1, 1] or [0, 1].

$$\bar{X}^{(j)} = \frac{x^{(j)} - min^{(j)}}{max^{(j)} - min(j)}$$

Asst.Prof.Dr. Boran Şekeroğlu

Feature Engineering – Data Normalization

- It is not a strict requirement.
- However, in practice, it can lead to an increased speed of learning.

Feature Engineering – Dealing with missing features

- In some examples, values of some features can be missing.
- That often happens when the dataset was handcrafted.

- How Do We Solve It?
  - Removing the examples with missing features from the dataset (if it is big enough).
  - Using a learning algorithm that can deal with missing feature values.
  - Using a data imputation technique.

Feature Engineering – Data Imputation Techniques

- Replace it by taking the average of all features:

ORIGINAL	NEW
A	A
0.5	0.5
0.6	0.6
0.55	0.55
0.4	0.4
NULL	0.542
0.66	0.66

Asst.Prof.Dr. Boran Şekeroğlu

Feature Engineering – Data Imputation Techniques

- Replace it by a value outside the normal range of values:
  - Ex. normal range → [0,1], replace it by 2.
  - Why? the learning algorithm will learn what is it better to do when the feature has a value significantly different from other values.

ORIGINAL	NEW
A	A
0.5	0.5
0.6	0.6
0.55	0.55
0.4	0.4
NULL	2
0.66	0.66

Feature Engineering – Data Imputation Techniques

- Replace it by a value in the middle of the range:
  - Ex. normal range → [0,1], replace it by 0.5.
  - Why? It will not significantly affect the prediction.

ORIGINAL	NEW
A	A
0.5	0.5
0.6	0.6
0.55	0.55
0.4	0.4
NULL	0.5
0.66	0.66

Feature Engineering – Data Imputation Techniques

- Use missing value as target for regression problem to predict.
- **Before you start working on the learning problem, you cannot tell which data imputation technique will work the best.**
- **Try several techniques, build several models and select the one that works the best.**

Feature Engineering – Learning Algorithm Selection

- Difficult task…

Deep

NN

SVM

Decision Trees

Linear Regression

Feature Engineering – Learning Algorithm Selection

- Explainability…
  - NN→ better results → very hard to understand
  - kNN, linear regression, decision trees → not always the most accurate →prediction is very straightforward
- Number of features and examples…
  - NN or Deep NN can handle huge data at the same time.
  - Others like SVM can not.
- Nonlinearity of the data…
  - Linearly separable? YES? → SVM, Logistic Regression, Linear Regression…
  - NO? NN or deep neural networks…

Feature Engineering – Learning Algorithm Selection

- Training speed
  - Deep Learning → Too Slow
  - NN → Slow
  - Simple algorithms like logistic and linear regression as well as decision tree learning → Much faster.

Feature Engineering

- **THREE SETS**
  - **Training Set :** the biggest one, and you use it to build the model.
  - **Validation Set :** 1) choose the learning algorithm and 2) find the best values of hyperparameters.
  - **Test Set :** to assess the model before delivering it to the client or putting it in production.

Feature Engineering

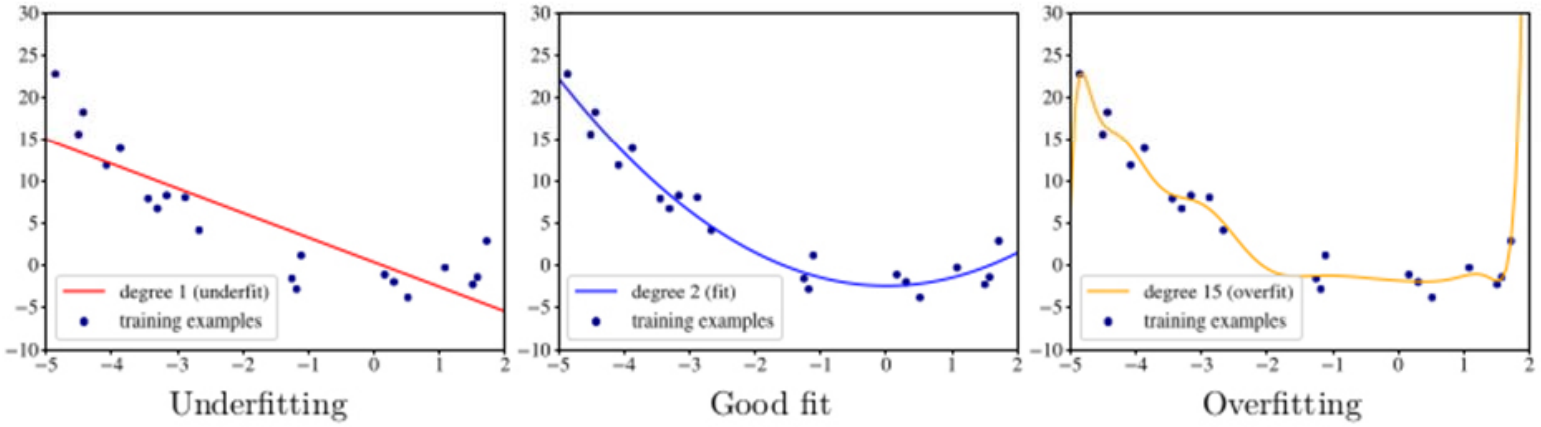
- **UNDERFITTING & OVERFITTING**

  - If the model makes many mistakes on the training data, we say that the model has a high bias or that the model underfits.

  - The model that overfits predicts very well the training data but poorly the data from at least one of the two hold-out sets.

Feature Engineering

- **UNDERFITTING & OVERFITTING**



Asst.Prof.Dr. Boran Şekeroğlu

Model Performance Assesment

- In Prediction:
  - **Mean Squared Error (MSE):** average of the square of the errors.

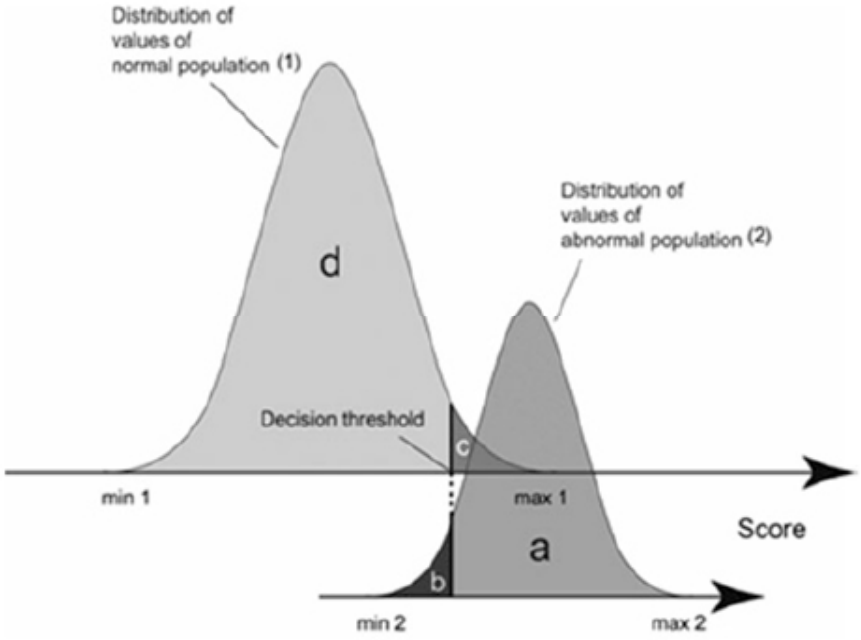
$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Model Performance Assesment

- In Prediction:
  - **R^2 Score:** related to the MSE… but not the same.
    - Total variance explained by model / total variance
    - Variance: measure of how far observed values differ from the average of predicted values.

Model Performance Assesment

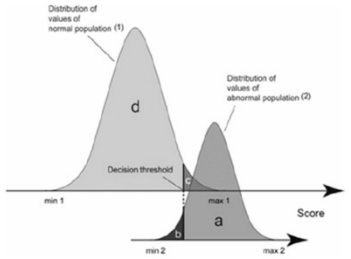
- In Classification:
  - EXAMPLE:



Asst.Prof.Dr. Boran Şekeroğlu

Model Performance Assesment

- True Positive, True Negative, False Positive, False Negative
  - ***a*** tests positive and actually has the disease: True Positive (TP)
  - ***b*** tests negative although he/she actually has the disease: False Negative (FN)
  - ***c*** tests positive but does not have the disease: False Positive (FP)
  - ***d*** tests negative and does not have the disease: True Negative (TN)



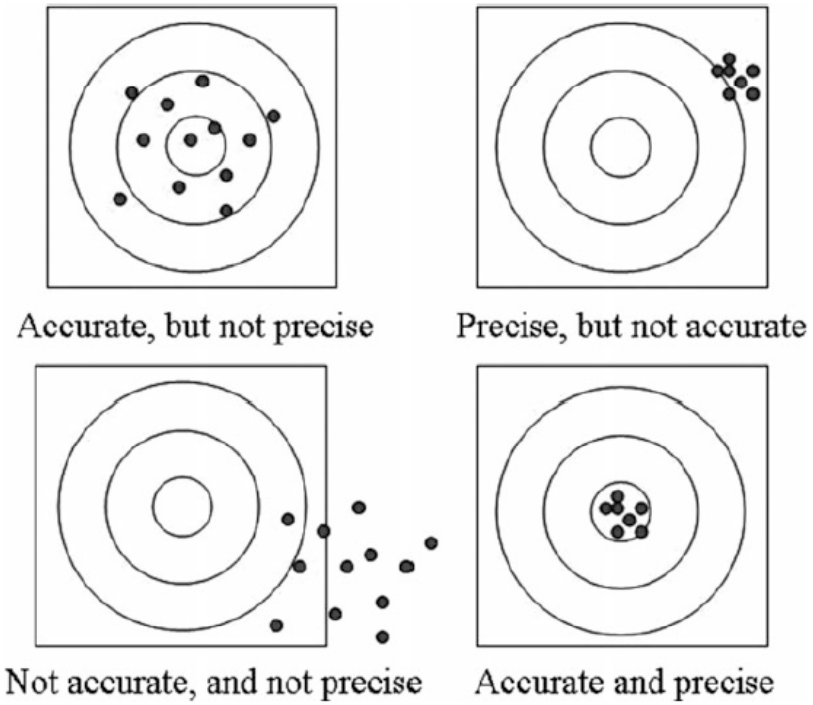
Asst.Prof.Dr. Boran Şekeroğlu

Model Performance Assesment

- Confusion Matrix

	B	$\bar{B}$
A	a (TP)	b (FN)
$\bar{A}$	c (FP)	d (TN)

Model Performance Assesment – Accuracy & Precision



Asst.Prof.Dr. Boran Şekeroğlu

Model Performance Assesment

- In Classification:
  - **Accuracy:** general success rate…

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}.$$

  - **Precision:**

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}.$$

Model Performance Assesment

- In Classification:
  - **Sensitivity:** propability of test identifying those with the disease given that they have the disease.

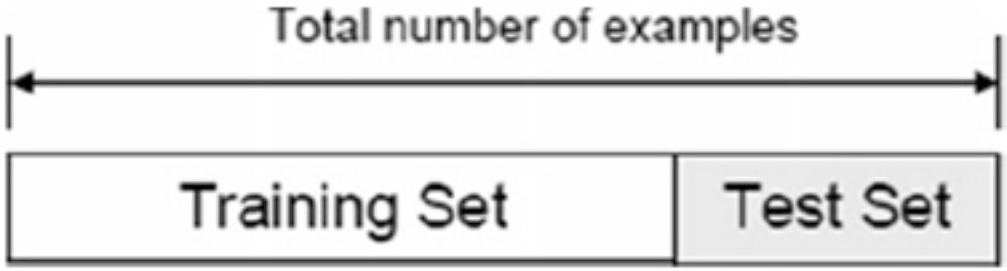
$$Sensitivity = \frac{TP}{TP + FN}$$

  - **Specificity:** propability of test identifying those free of the disease given that they do not have the disease.

$$Specificity = \frac{TN}{TN + FP}$$

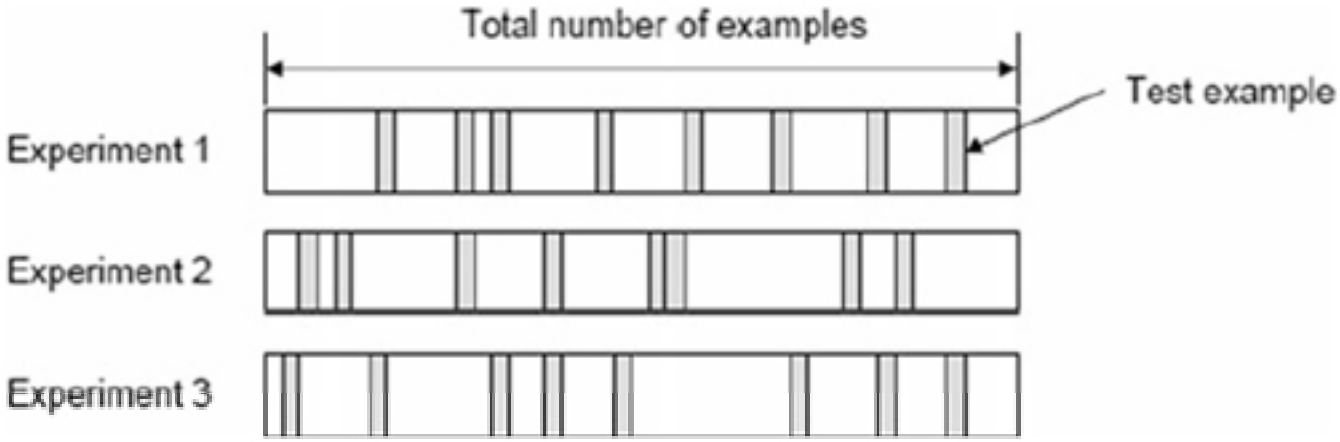
Model Performance Assesment Cross-validation and Resampling

- The Holdout Method
  - Simplest kind…
  - The dataset is separated into two sets, called the training set and the test set.
  - Not all the data is used for training…



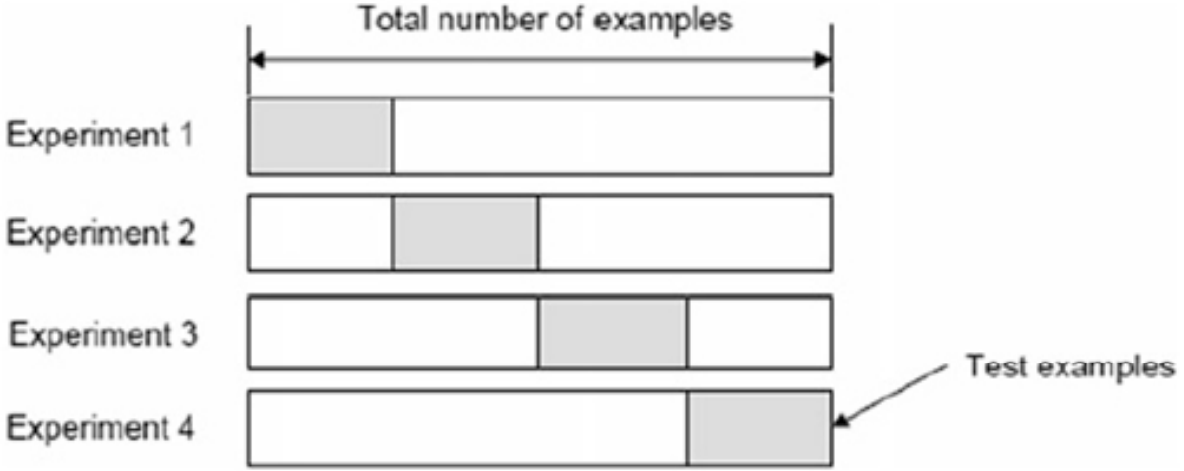
Model Performance Assesment Cross-validation and Resampling

- Random Sub-sampling
  - the holdout method is repeated several times.



Model Performance Assesment Cross-validation and Resampling

- k-fold Cross-validation
  - The dataset is divided into k equal-sized subsets
  - One for testing and the for training
  - Repeated k-times – each subset is used for testing exactly once
  - The total error is obtained by averaging the errors for all the runs



Machine Learning Techniques

- **Some Prediction Algorithms:**
  - Support Vector Regression
  - Linear Regression
  - Decision Tree Regressor

- **Some Classification Algorithms:**
  - Logistic Regression
  - Support Vector Machines
  - Decision Trees
  - K-nearest neighbors
  - Naive-Bayes

Machine Learning Techniques

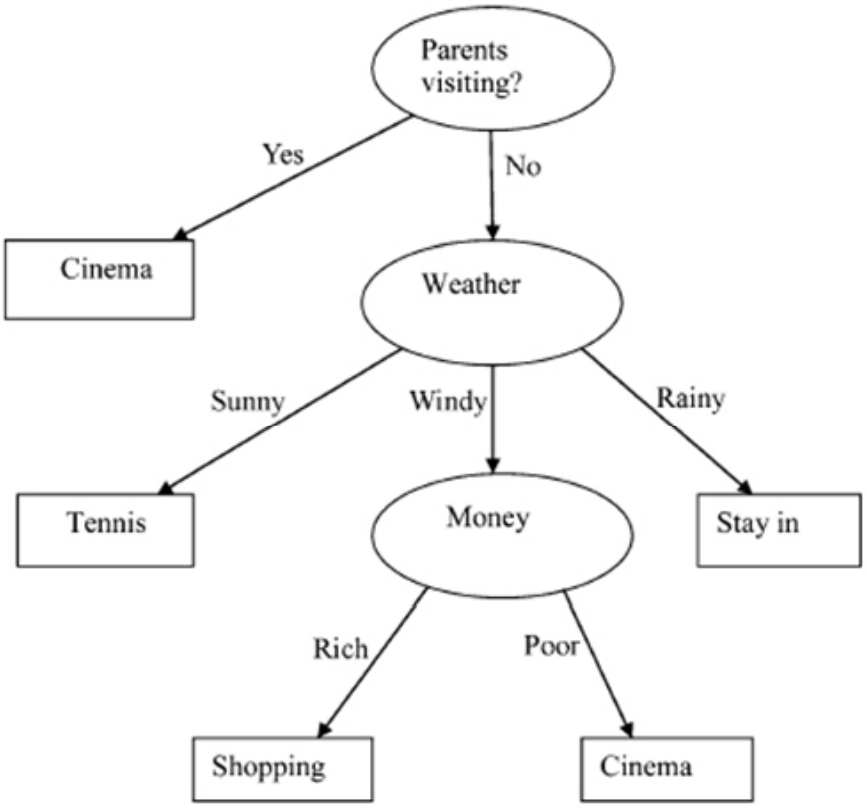
- **Prediction & Classification**
  - Neural Networks
  - Deep Learning

Machine Learning Techniques

- **Decision Trees**
  - simple classifier in the form of a hierarchical tree structure
  - supervised classification
  - *divide-and-conquer* strategy
  - Starting node is known as the **root node** which is considered the parent of every other node
  - Successive decision nodes are visited until a terminal or **leaf node** is reached, where the **class** is assigned

Machine Learning Techniques

- **Decision Trees**



Asst.Prof.Dr. Boran Şekeroğlu

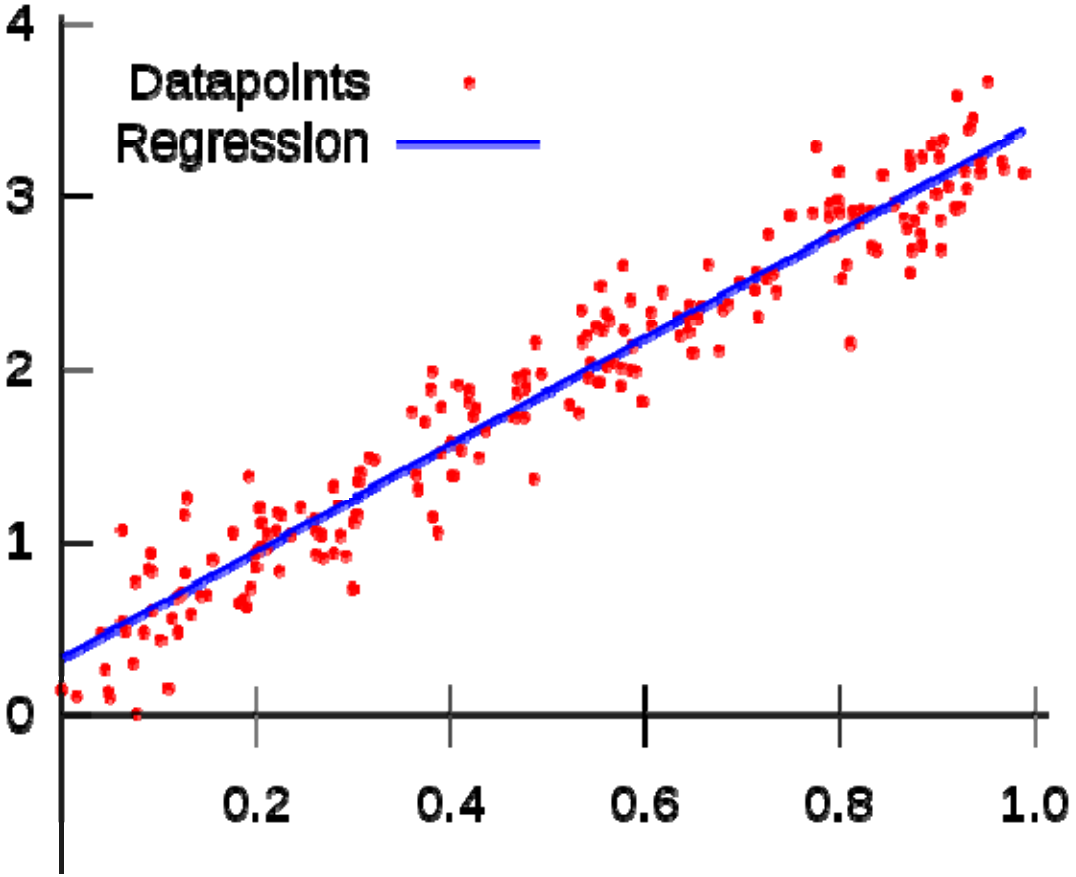
Machine Learning Techniques

- **Decision Trees**
  - Advantages:
    - can be used with nonmetric/categorical data
    - interpretable
    - very fast and require very little computation
  - Disadvantages:
    - Exponentially many decision trees that can be constructed from a given set of features
    - Some of the trees will be more accurate than others
    - Finding the optimal tree is not computationally feasible

Machine Learning Techniques

- **Linear Regression**
  - finding relationship between two continuous variables.
  - One is predictor (or independent variable) and other is response (or dependent variable).
  - Idea is to obtain a line that best fits the data.
  - The best fit line: the one for which the total prediction error are as small as possible.
  - Efficient for linearly separable prediction problems…

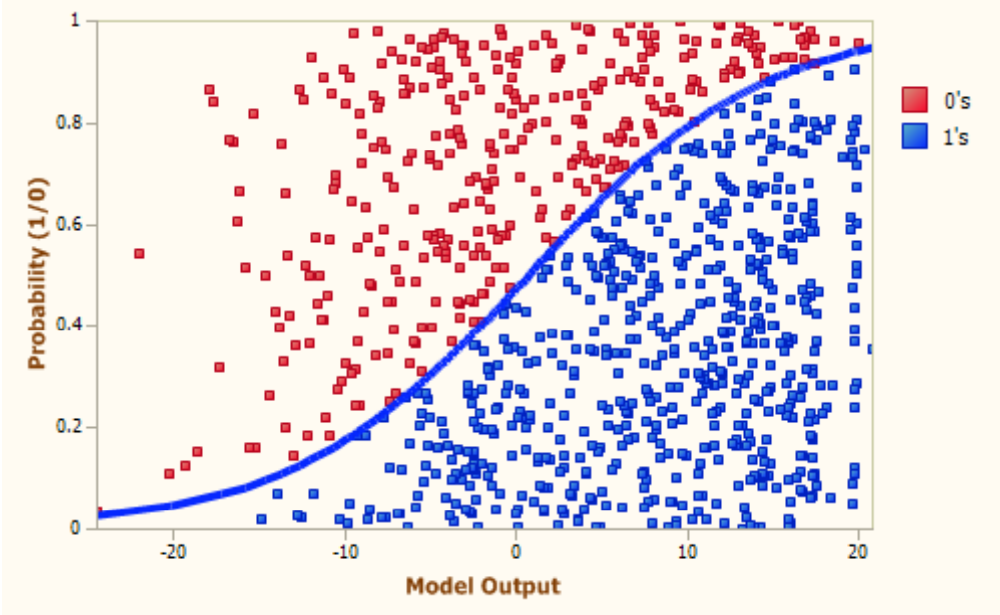
Machine Learning Techniques



Asst.Prof.Dr. Boran Şekeroğlu

Machine Learning Techniques

- Logistic Regression
  - Uses Logistic (Sigmoid function)



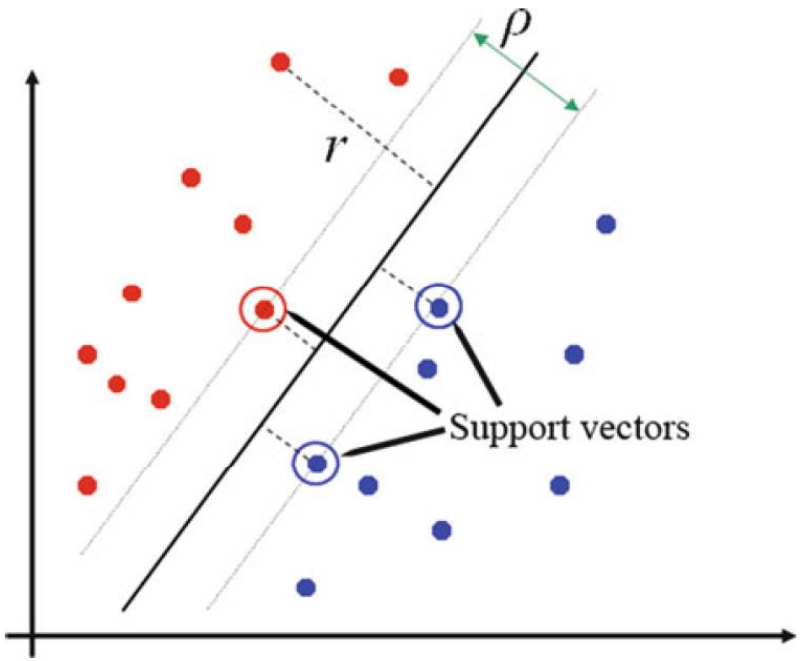
Asst.Prof.Dr. Boran Şekeroğlu

Machine Learning Techniques

- **Support Vector Machines (Kernel Machines)**
  - became popular when it was applied to a hand-writing recognition task.
  - For linearly separable problems, it finds the optimal separating hyperplane by maximizing the margin, the perpendicular distance across the hyperplane to closest instances (the support vectors).
  - SVMs are basically two-class classifiers.
  - Typical kernel functions are linear, polynomials and Gaussians (radial basis functions).

Machine Learning Techniques

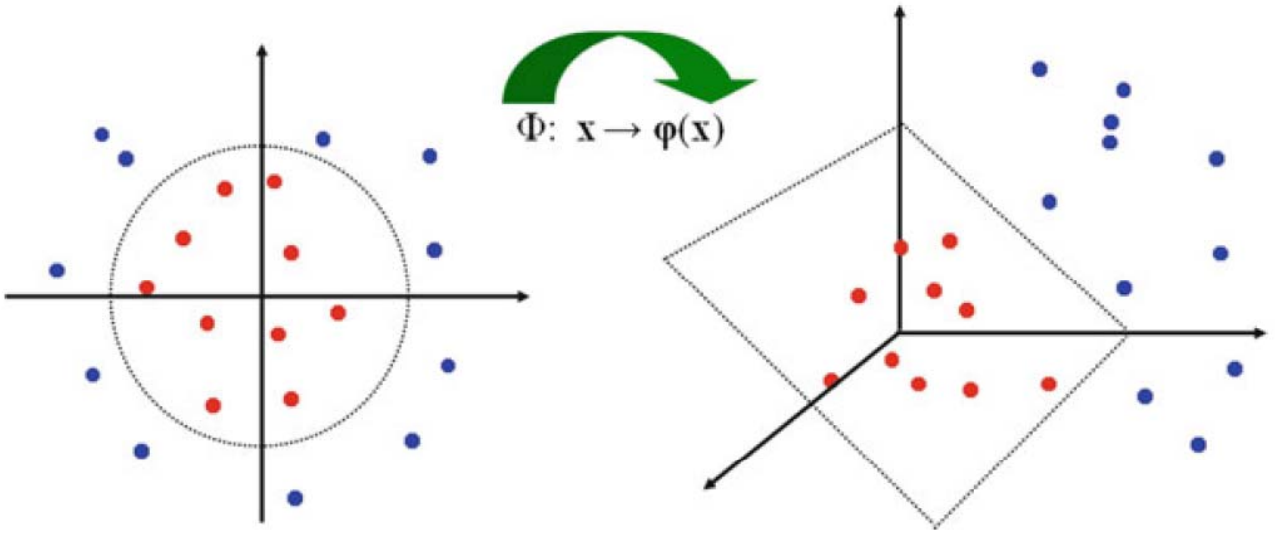
- **Support Vector Machines (Kernel Machines)**



Asst.Prof.Dr. Boran Şekeroğlu

Machine Learning Techniques

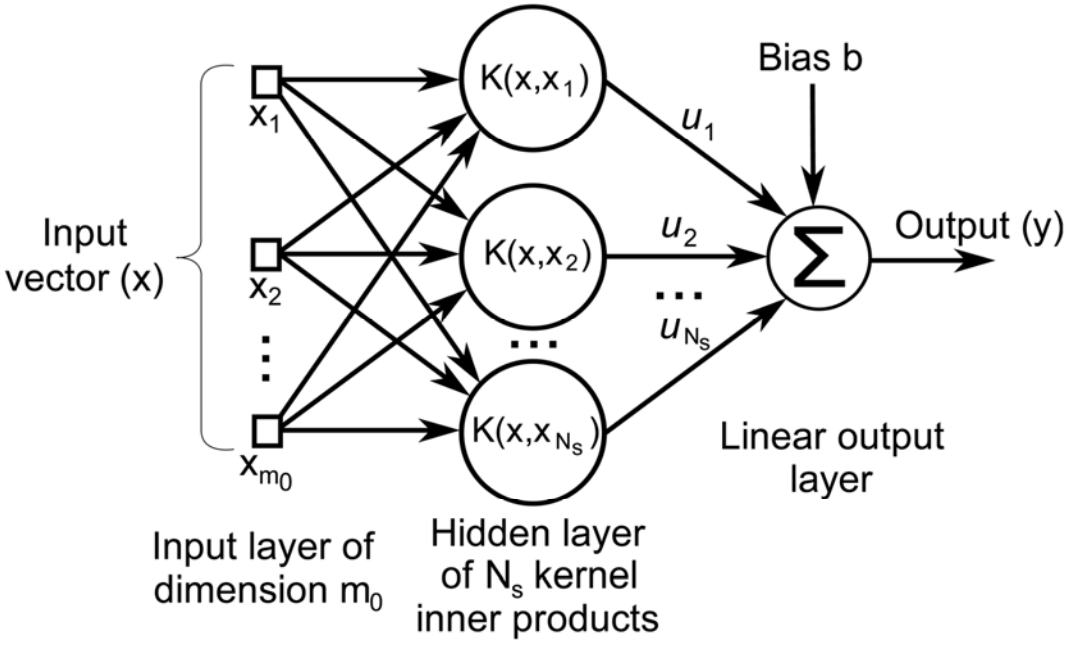
- **Support Vector Machines (Kernel Machines)**



Asst.Prof.Dr. Boran Şekeroğlu

Machine Learning Techniques

- **Support Vector Machines (Kernel Machines)**

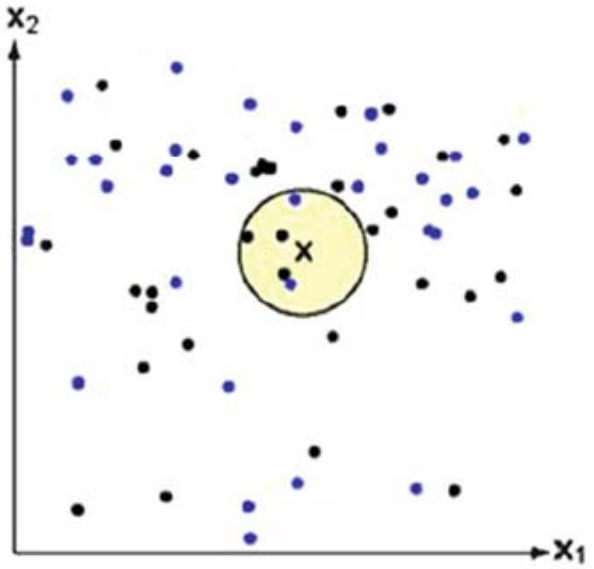


Machine Learning Techniques

- **k-Nearest Neighbor**
  - The k-NN process starts at the test point and grows a region until it encloses k training samples and it labels the test point **x** by majority vote of these samples.
  - For two classes, the value of k should be odd to avoid a tie.
  - For more than two classes, k being odd is insufficient to avoid a tie [3 classes→ k=5, tie (2,2,1) in some cases].

Machine Learning Techniques

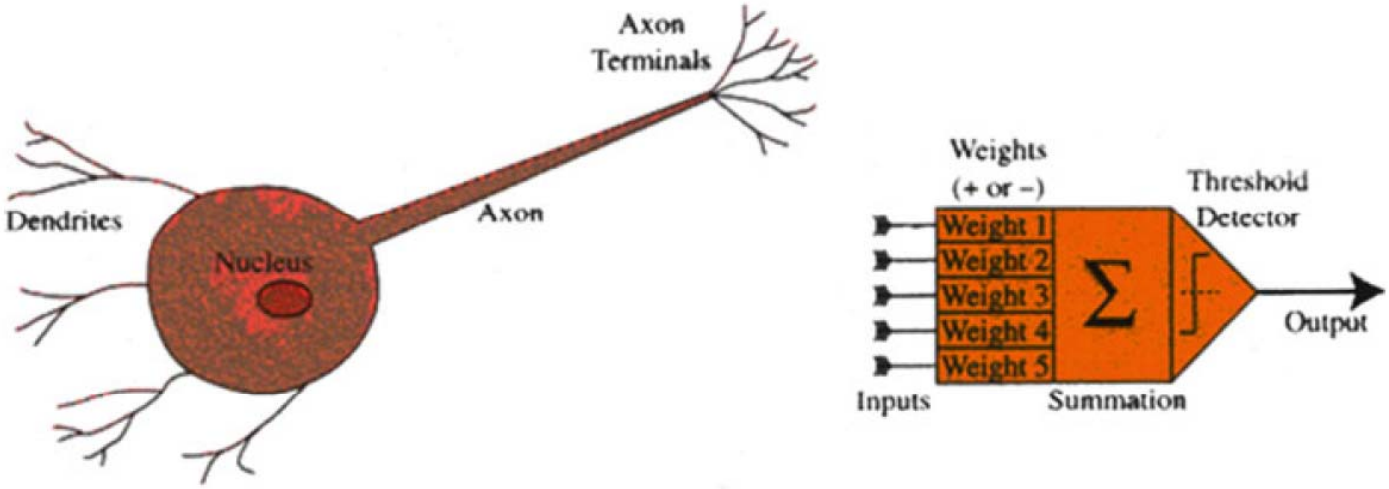
- **k-Nearest Neighbor**



Asst.Prof.Dr. Boran Şekeroğlu

Machine Learning Techniques

- **(Artificial) Neural Networks**
  - models take their motivation from the human nervous system.



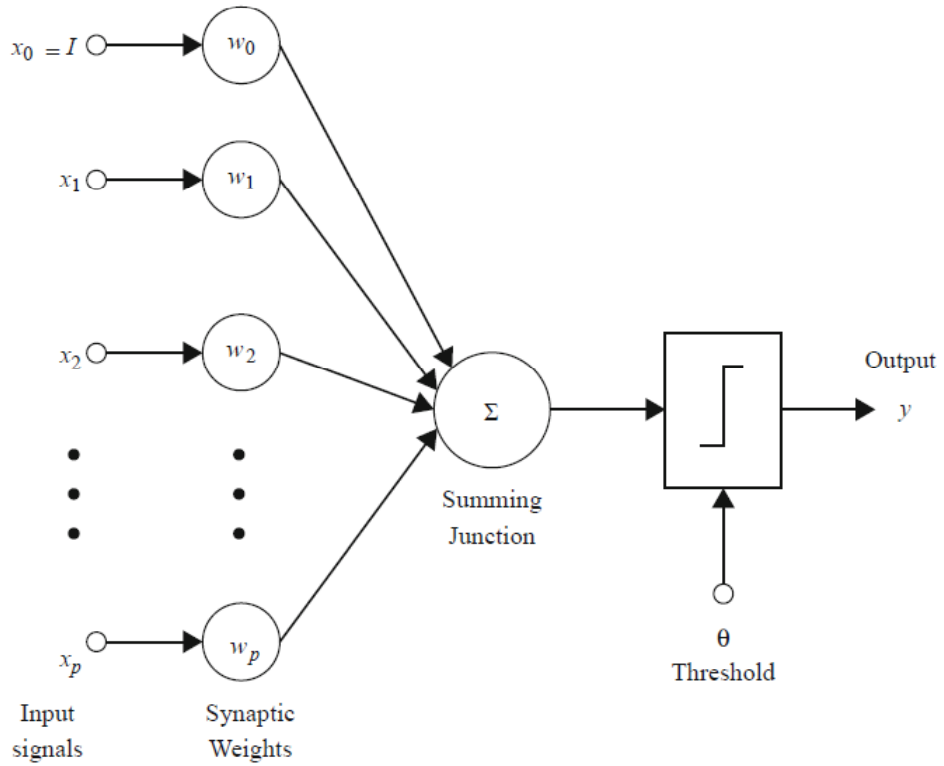
Asst.Prof.Dr. Boran Şekeroğlu

Machine Learning Techniques

- ANN is an adaptive system, i.e., parameters can be changed during operation (training) to suit the problem.
- They can be used in a wide variety of classification tasks, e.g., character recognition, speech recognition, fraud detection, medical diagnosis etc.

Machine Learning Techniques

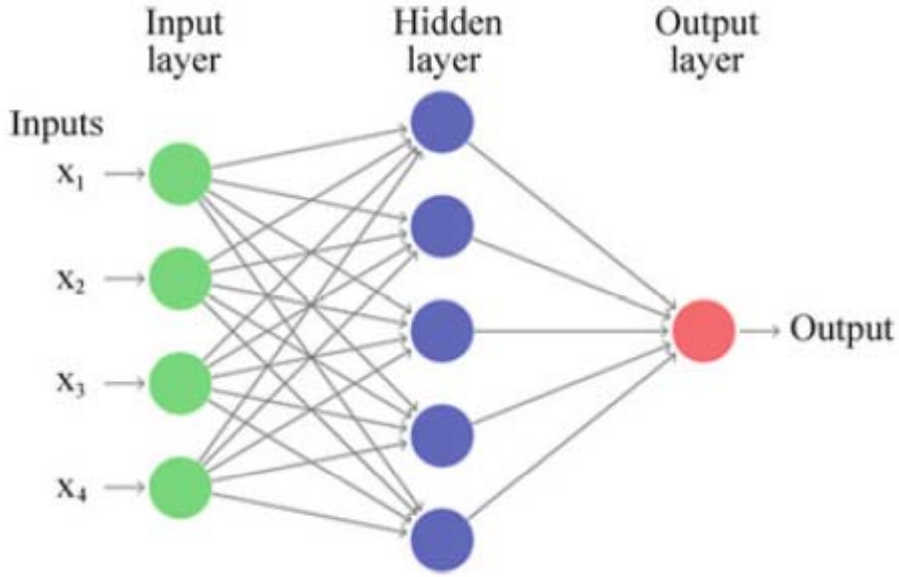
- McCullough and Pitts Neuron



Asst.Prof.Dr. Boran Şekeroğlu

Machine Learning Techniques

- Neural networks learn using an algorithm called *backpropagation*.

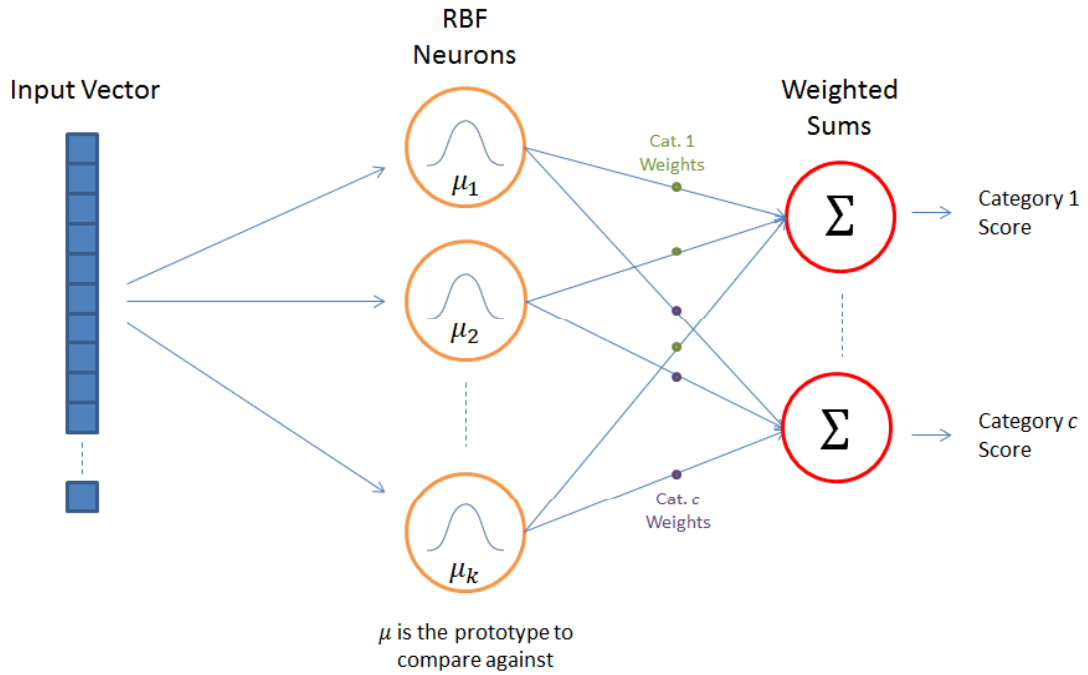


Machine Learning Techniques

- **ANN Algorithms:**
  - **Backpropagation Learning Algorithm**
    - One of the most famous algorithm for both classification and prediction problems.
    - Easy in implementation, but difficult in determining parameters
    - Uses gradient-descent algorithm to update weights
    - Faster than others during training
  - **Radial-Basis Function Neural Network**
    - Similar to backpropagation
    - Uses radial-basis function in hidden layer
    - Generally produces more accurate results than BP
    - Needs more training time

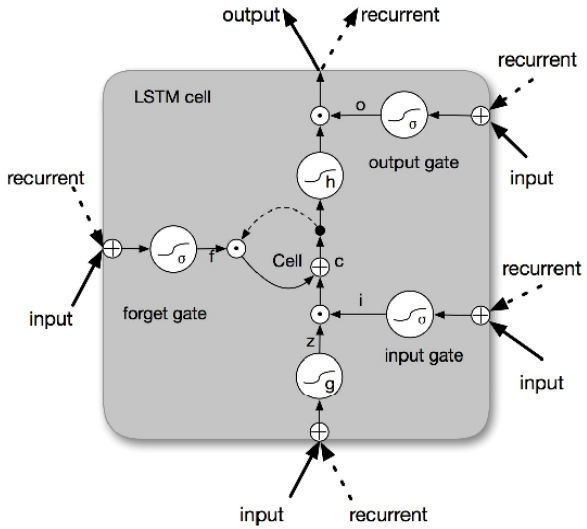
Machine Learning Techniques

- **ANN Algorithms:**
  - **Radial-Basis Function Neural Network**



Machine Learning Techniques

- **ANN Algorithms:**
  - **Long-Short Term Memory (LSTM) Neural Network**
    - One of the most accurate NN algorithm for prediction problems
    - difficulty in implementation,
    - Needs huge computational time during training

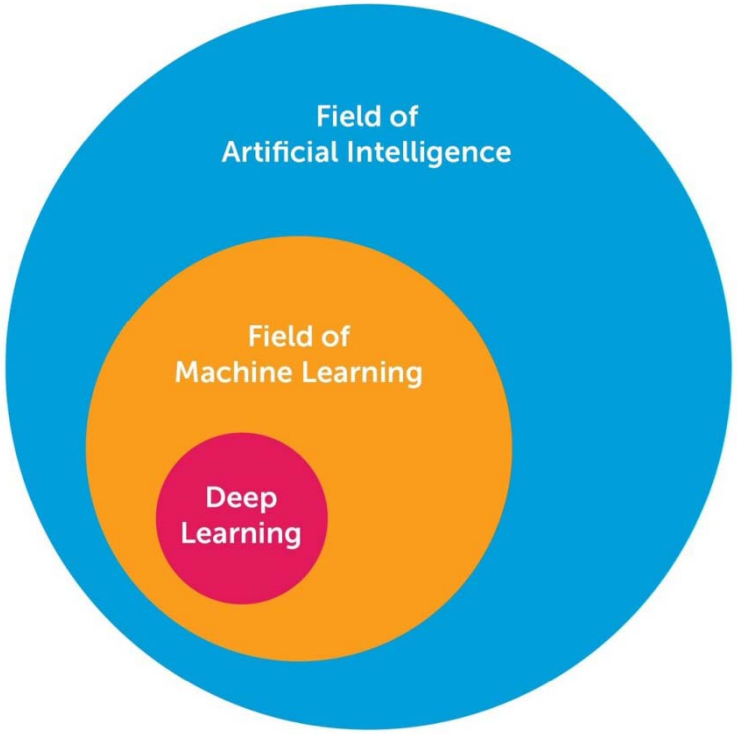


Machine Learning Techniques

- **ANN Algorithms:**
  - ANN Algorithms have several modified and improved versions for specific problems.
  - ANN produces accurate results for non-linear problems…
  - Parameter selection is the main challenge…

Machine Learning Techniques

- **Deep Learning = Deep Structured Learning = Hierarchical Learning**



MACHINE LEARNING TECHNIQUES

- The "deep" : the number of layers through which the data is transformed.
- Most modern Deep Learning methods are based on ANN.
- **Deep Learning:**
  - More than two hidden layers…
  - *Convolutional Neural Networks:* is designed to recognize images by having convolutions inside, which see the edges of an object recognized on the **image**.
  - *Recurrent Neural Network:* is basically a standard neural network that has been extended across time by having edges which feed into the next time step instead of into the next layer in the same time step. Efficient for **speech and text recognition**.

Machine Learning Techniques

- *Deep Autoencoder:* unsupervised… designed for feature extraction.
- *Deep Boltzman Machine:* undirected connections between layers… both supervised and unsupervised.

End of Session -I-