



Machine Learning

1. week

- Terminology
- Supervised – Unsupervised Learning
- Data Preparation
- Cross Validation
- Overfitting



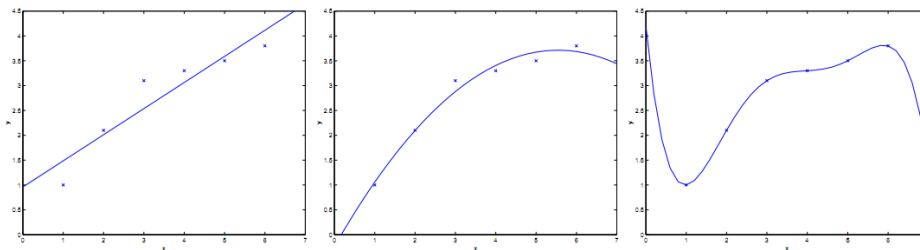
What is Machine Learning ?

Machine Learning is common name of algorithms which can model a problem according to its data.

There are many approaches. A part of these approaches is on prediction and estimation, other part is on classification.

Machine Learning Methods

Machine learning methods can be different to each other according to their approach to problem and therefore may have different success in different problems.



Umut ORHAN, PhD.

3

Machine Learning Terms

- Prediction: It is used in case desired outputs have to be quantitative (continuous).
- Classification: It is used in case desired outputs have to be qualitative (discrete).

Umut ORHAN, PhD.

4

Prediction And Estimation

Because they have similar meaning, the usages are confused in literature.

But in statistics, while estimate term is used for determination of the model, prediction is a computation of an unknown value of a random variable by using an estimated equation.

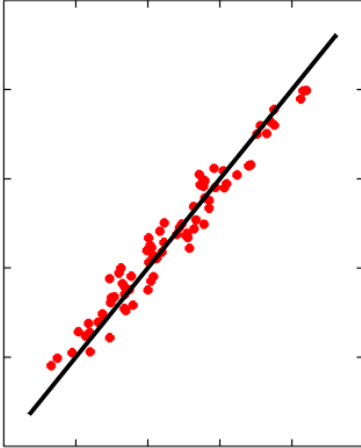
Umut ORHAN, PhD. 5

Prediction And Estimation

For example, a linear curve function is determined on red points.

"Estimation" term is used to determine equation represent relation between x and y ,

"Prediction" term should be preferred to compute y values corresponding to an x value.



Umut ORHAN, PhD. 6



Estimation Example

With regard to a problem, let following dataset be given. The goal is to produce an equation which can compute y value corresponding any x input.

It is easy to see that the solution of this simple problem is $y = 3 * x$.

X	Y
3	9
5	15
8	24
10	30
19	57
24	72
27	81
31	93
38	114
43	129

Umut ORHAN, PhD.

7



Prediction Example

According to the equation of found solution $y = 3 * x$, if y value is desired for $x = 50$ input, then it is possible to easily compute $y = 150$ value.

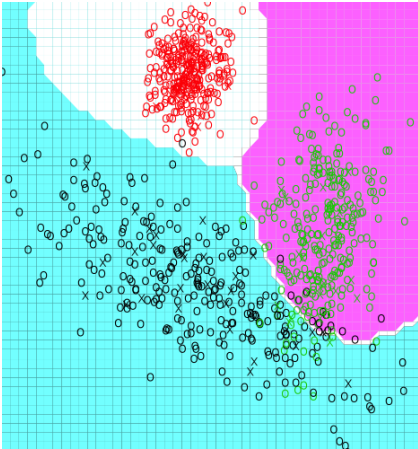
X	Y
3	9
5	15
8	24
10	30
19	57
24	72
27	81
31	93
38	114
43	129

Umut ORHAN, PhD.

8

Classification

The goal is to divide the whole space of problem into a certain number of classes. In the image at the right, each color represents a class. By means of classification techniques, all space, even areas without any data, can be painted.



Umut ORHAN, PhD.

9

Classification Example


Because Y values are all discrete, this is a classification problem. Classification equation can be written easily by using a threshold value as below.

$$Y = \begin{cases} 0, & X < 20 \\ 1, & X \geq 20 \end{cases}$$

X	Y
2	0
5	0
9	0
13	0
19	0
20	1
27	1
33	1
39	1
47	1

Umut ORHAN, PhD.

10




Supervised vs. Unsupervised

According to dataset, learning is performed in two different ways.

- Supervised Learning: Data is organized as input vs. output parameters. The aim is to compute output values by using inputs in minimum errors.
- Unsupervised Learning: The aim to discover hidden groups in data without any output information.

Umut ORHAN, PhD. 11




Reinforcement Learning

Sometimes, supervisor does not give expected result directly to learning machine. But for produced results, partial supports are send to system as "true / false". This learning method is called as reinforcement learning.

Boltzmann machine, LVQ and genetic algorithm can be considered as examples.

Umut ORHAN, PhD. 12




Clustering

All classification and estimation methods can be considered as supervised learning methods.

Clustering methods are described under unsupervised learning title.

But what is clustering in detail?

Umut ORHAN, PhD. 13



Clustering

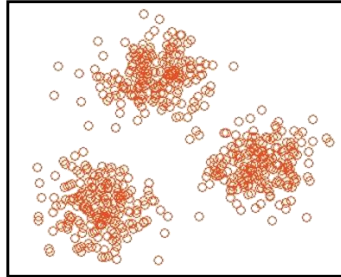
Clustering dealing with unlabeled data is the process of organizing similar objects into the same groups and dissimilar objects into different groups.

In clustering literature, similarity term is used for opposite sense of distance.

Umut ORHAN, PhD. 14

Clustering

Because of their similarity, closer samples to each other are placed into the same cluster. Likewise, distant samples are located in different clusters. The number of clusters is usually provided by experts.

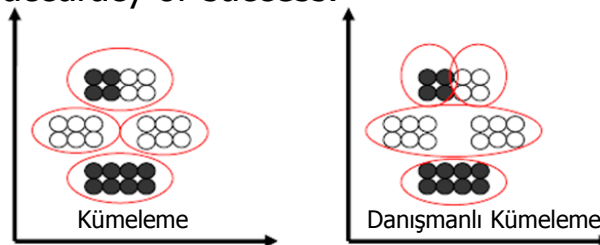


Umut ORHAN, PhD.

15


Supervised Clustering

It is a supervised learning that can use class information and sample similarities at the same time. On other hands, clustering before classical classification usually increases the classification accuracy or success.



Umut ORHAN, PhD.

16




Notation

In our studies, notations at below are preferred.

- D for desired classes,
- Y for estimation outputs,
- X for input data,
- X_j for each feature of input,
- x_i for each example,
- \mathbf{X} for all of the input data set


Umut ORHAN, PhD. 17



Learning Schedule

- Online learning: If the learning process should be sustained continuously, online learning is used.
- Offline learning: At first, the system is trained, then it loaded, and started. This kind of systems does not have to run continuously.

Umut ORHAN, PhD. 18




Learning Rules

Although there are many training algorithms proposed, according to learning rule background, learning algorithms can be divided into four groups:

- Hebb
- Delta
- Hopfield
- Kohonen

Umut ORHAN, PhD. 19




Hebb Learning Rule

It is the first learning rule developed in 1949.

The rule is based on principle of "a cell affects its neighbors".

By improving this rule, different learning rules have been developed.

Umut ORHAN, PhD. 20




Delta Learning Rule

Squared difference between the desired and the calculated results is the error of the system.

In order to reduce this error, connections between cells is continuously updated.

Multilayered perceptron networks are trained in accordance with this rule.

Umut ORHAN, PhD. 21



Hopfield Learning Rule

When the desired result is the same as the calculated one, connection between related cells is strengthened in a specific ratio.

Otherwise the connection is weakened.

Recurrent Elman networks are trained with this rule.

Umut ORHAN, PhD. 22

Kohonen Learning Rule

In this unsupervised model, cells are in the race. The cell that produces the greatest result wins the race. All connections of winner cell are strengthened.

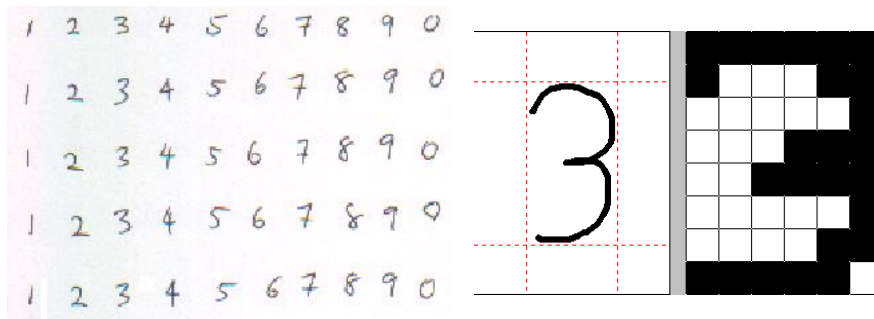
ART (Adaptive Resonance Theory) and SOM (Self Organizing Map) developed by Kohonen are examples for this rule.

Umut ORHAN, PhD.

23

Data Preparation

Before starting learning process, a dataset that can represent whole problem space should be prepared.

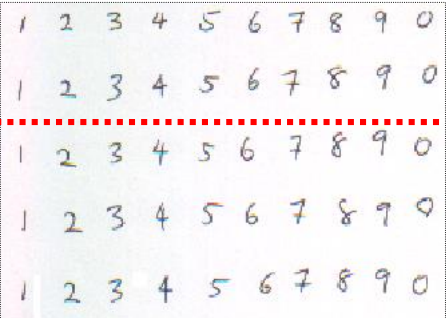


Umut ORHAN, PhD.

24

Data Preparation

Data prepared for the solution is divided into two sets because it is used in both train and test processes.



Umut ORHAN, PhD. 25


Validation

The main aim of machine learning studies is to provide that the system trained from a dataset can answer any unknown question in the same problem.

Therefore the limited data should be used in both training and testing the system.

The methods known as cross-validation are successful in this subject.


Umut ORHAN, PhD. 26



Cross Validation

This proposed method is based on the basic principles are a few validity. But all has the same basic logic. The current data set to measure the success of the system is divided into two parts. One for training (train sets) and the other is never seen to represent possible examples of the system (test set) is used. The system learns the training set with the selected training algorithm. The success of the trained system is then calculated on the test set.

Umut ORHAN, PhD. 27



Cross Validation

Three types of cross-validation method is proposed:

- Random sampling
- K-fold
- Leave one out

Umut ORHAN, PhD. 28

Random Sampling

1

2

3

4

Umut ORHAN, PhD.

29

K-Fold

DATASET

DATASET

DATASET


DATASET

DATASET

Red folds show test-set, Blue ones are train-set.

Umut ORHAN, PhD.

30




Leave One Out

It is a special application of K-fold ($K = N$).

For a dataset with N samples, if we choose number of folds (K) as number of samples (N), then...

K-fold runs as Leave One Out method.

Umut ORHAN, PhD. 31



Overfitting

All iterational learning machines must be stopped at the right time.

Otherwise the system starts to memorize examples in the training data, and this decreases the prediction ability of the system for unknown samples.

This kind of excessive training is called as Overfitting.

Umut ORHAN, PhD. 32