

# Machine Learning and Classification

A. Zemmari

LaBRI - University of Bordeaux

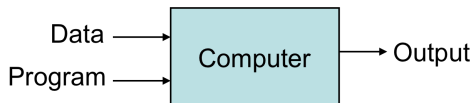
Nov. 28, 2017

# What is Machine Learning

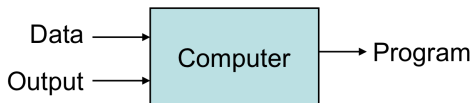
- ▶ "The goal of machine learning is to build computer systems that can adapt and learn from their experience." - Tom Dietterich
- ▶ To learn: to acquire knowledge by study, experience or being taught.

# What is Machine Learning

- ▶ Traditional Programming



- ▶ Machine Learning



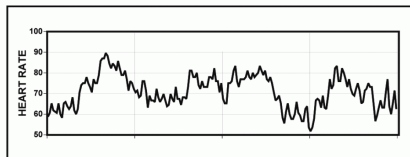
# Magic?

No, more like gardening:

- ▶ Seeds = Algorithms
- ▶ Nutrients = Data
- ▶ Gardener = You
- ▶ Plants = Programs

## A Few Examples

- ▶ Given 20 years of clinical data, will this patient have a second heart attack in the next 5 years?



## A Few Examples

- ▶ Given 20 years of clinical data, will this patient have a second heart attack in the next 5 years?
- ▶ What price for this stock, 6 months from now?



15.89	+5.31%	▲	251.23	320,000
7.34	+5.97%	▲	100.00	120,000
1.89	+2.13%	▲	564.23	300,000
1.45	+6.43%	▲	765.90	600,000
1.67	-11.6%	▼	120.34	380,000
1.54	+23.1%	▲	893.23	120,000
1.19	+5.56%	▲	128.98	320,000
1.08	-3.67%	▼	432.12	750,000
1.01	+11.3%	▲	765.23	150,000
1.00	+2.54%	▲	432.24	120,000

## A Few Examples

- ▶ Given 20 years of clinical data, will this patient have a second heart attack in the next 5 years?
- ▶ What price for this stock, 6 months from now?
- ▶ Is this handwritten number a 7?



# A Few Examples

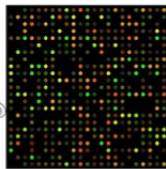
- ▶ Given 20 years of clinical data, will this patient have a second heart attack in the next 5 years?
- ▶ What price for this stock, 6 months from now?
- ▶ Is this handwritten number a 7?
- ▶ Is this e-mail I just received a spam ?





# A Few Examples

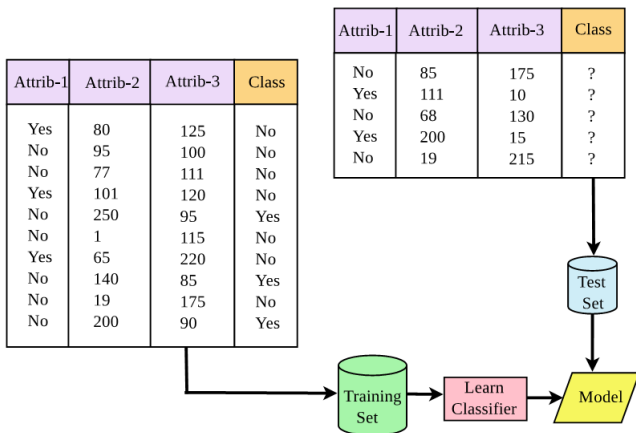
- ▶ Given 20 years of clinical data, will this patient have a second heart attack in the next 5 years?
- ▶ What price for this stock, 6 months from now?
- ▶ Is this handwritten number a 7?
- ▶ Is this e-mail I just received a spam ?
- ▶ Can I cluster together different customers? words? genes?



# Learning Models

- ▶ **Supervised Learning:** Assumes that the training examples are classified (**labelled by class labels**).
- ▶ The correct results are known in advance and are given as input to the model during the learning process.
- ▶ The construction of a precise training, validation and test set is crucial.
- ▶ **Predictive** data mining methods are supervised. They are used to induce models or theories (such as decision trees or rule sets) from class-labeled data, later used for prediction and classification.

# Supervised Learning



# Unsupervised Learning

- ▶ Unsupervised Learning Concern about the instances when class labels are unknown and model is provided with incomplete information.
- ▶ Can be used to cluster the input data in classes on the basis of their statistical properties.
- ▶ The labeling can be carried out even if the labels are only available for a small number of objects representative of the desired classes.

# Unsupervised Learning

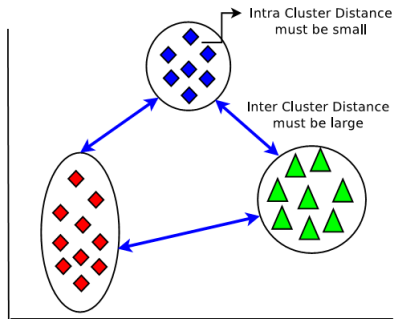


Figure: Unsupervised Learning

# Deep Learning

- ▶ What exactly is deep learning ?
- ▶ Why is it generally better than other methods on image, speech and certain other types of data?

The short answers

- ▶ Deep Learning means using a neural network with several layers of nodes between input and output.
- ▶ The series of layers between input & output do feature identification and processing in a series of stages, just as our brains seem to.

# Deep Learning

- ▶ What exactly is deep learning ?
- ▶ Why is it generally better than other methods on image, speech and certain other types of data?

The short answers

- ▶ Deep Learning means using a neural network with several layers of nodes between input and output.
- ▶ The series of layers between input & output do feature identification and processing in a series of stages, just as our brains seem to.

# Deep Learning

- ▶ What exactly is deep learning ?
- ▶ Why is it generally better than other methods on image, speech and certain other types of data?

## The short answers

- ▶ Deep Learning means using a neural network with several layers of nodes between input and output.
- ▶ The series of layers between input & output do feature identification and processing in a series of stages, just as our brains seem to.



# Deep Learning

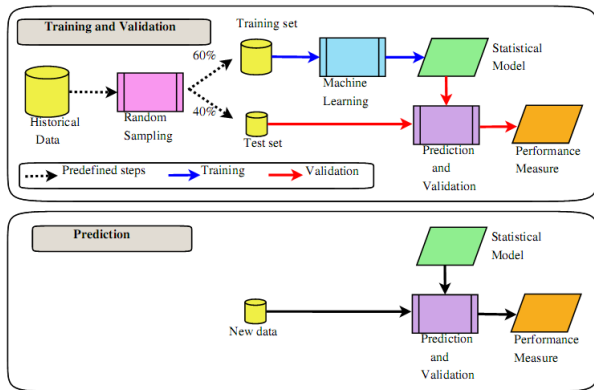
- ▶ What exactly is deep learning ?
- ▶ Why is it generally better than other methods on image, speech and certain other types of data?

## The short answers

- ▶ Deep Learning means using a neural network with several layers of nodes between input and output.
- ▶ The series of layers between input & output do feature identification and processing in a series of stages, just as our brains seem to.

In this presentation: we focus on supervised machine learning & deep learning.

# General Methodology



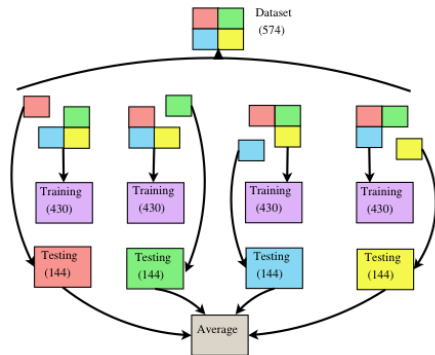
# How to evaluate a classification

- ▶ If the dataset is large enough:  
split the dataset samples into two subsets: **Training Set (60%)**  
and **Test Set (40%)**.
- ▶ otherwise: use cross-validation.

# How to evaluate a classification

- ▶ If the dataset is large enough:  
split the dataset samples into two subsets: **Training Set (60%)**  
and **Test Set (40%)**.
- ▶ otherwise: use cross-validation.

# How to evaluate a classification



# Evaluation Metrics

	Actual class (Observation)	
Predicted class (Expectation)	TP	FP
	FN	TN

- ▶ TP : True Positive
- ▶ FP : False Positive
- ▶ FN : False Negative
- ▶ TN : True Negative

Table: Evaluation parameters

$$\textit{Precision}(P) = \frac{TP}{TP + FP}$$

# Some Machine Learning Algorithms

- ▶ Naive Bayes classification.
- ▶ Perceptron and Multi-layers Perceptron.
- ▶ Support vector machines and kernel methods.
- ▶ Neural networks.
- ▶ Logistic Regression.
- ▶ Deep Learning: Convolutional Neural Networks.
- ▶ ...