Neural Networks and Learning Systems
TBMI 26, 2017

Lecture 1

Introduction

Magnus Borga
magnus.borga@liu.se

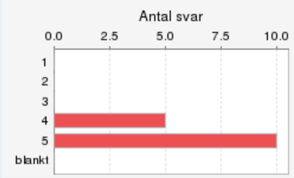
Course information

- All information will be available on Lisam

2

Course evaluation and development

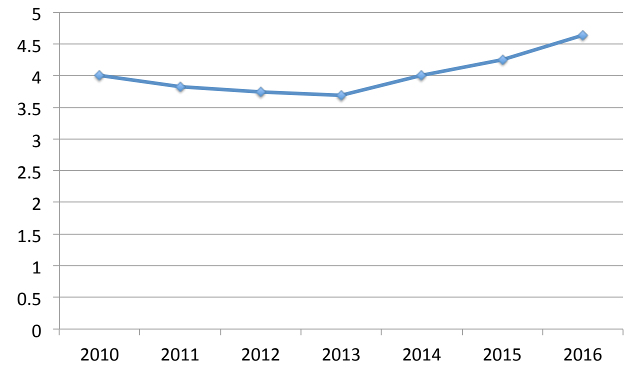
På en femgradig skala ger jag kursen sammanfattningsbetyget / On a scale 1-5 (5 being the best) I give the overall credit to this course



Svarsalternativ	Antal svar	Andel	Visa grupp
1			
2			
3			
4	5	33%	→
5	10	67%	→

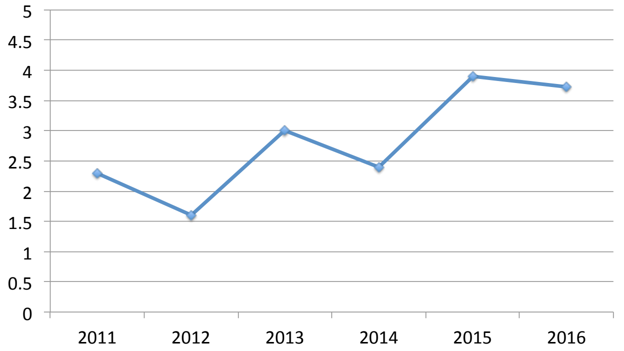
3

Overall credit over time



4

Literature credit



5

The Course - Overview

- 9 lectures
- 9 lessons
- 4 assignments
- 1 written exam

Must be completed

- Course language is English.

6

The Course - Lectures

PPT lectures, handouts on course page

1. Introduction (MB)
2. Supervised learning – Linear classification (OF)
3. Supervised learning – Neural networks (OF)
4. Supervised learning – Ensemble learning & Boosting methods (OF)
5. Supervised learning – Deep learning & summary (OF)
6. Unsupervised learning – Dimensionality reduction (MB)
7. Kernel methods (MB)
8. Clustering, Genetic algorithms (MB)
9. Reinforcement learning (MB)

7

The Course - Lessons

- One lesson after each lecture
- Pen & paper exercises
- Complementary presentations
- Preparations and help with assignments

8

The Course - Assignments

- 4 exercises/assignments
  - Pattern recognition using linear classifiers and neural networks
  - Face recognition in images using Boosting techniques
  - Unsupervised learning – PCA & LDA
  - Reinforcement learning
- Matlab
- In pairs or alone!
- Supervision time scheduled (“Laboration” in schedule) - two groups

- Hard deadline for written reports.
- No report for exercise 3 (PCA). Reported orally at one of the supervision time slots
- Late reports will not be corrected until the re-exam in June.

9

Course literature

- Lecture notes
- Recommended reading on the Lectures-page
- Exercise collection
- Assignments

- Additional link collection on course pages
  - Video lectures
  - Online books
  - Articles, tutorials
  - Applets

10

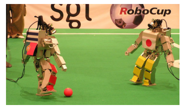
Prerequisites

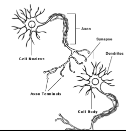
- Linear algebra
  - Vectors, scalar products, eigenvalues
- Multidimensional calculus
  - Gradients, partial derivatives
- Mathematical statistics
  - Mean, variance, covariance, correlation, Gaussian distribution,
- Programming
  - Some programming experience
  - Matlab knowledge helps a lot

11

Applications of machine learning

- Pattern and speech recognition
- Robots & autonomous systems
- Games
- Big data
- Evolutionary systems
- Expert systems & decision support
- Models of the brain

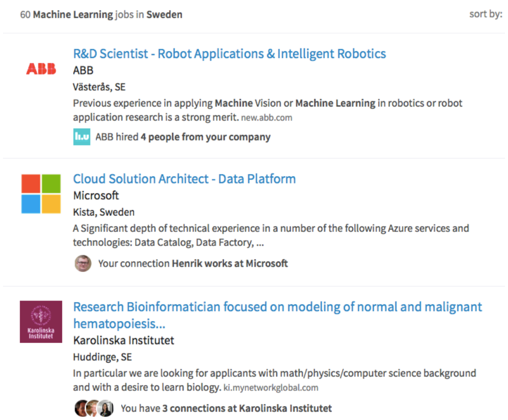




12

Jobs in Machine Learning

Jan 15, 2017


60 **Machine Learning** jobs in **Sweden** sort by:

R&D Scientist - Robot Applications & Intelligent Robotics
ABB
Västerås, SE
Previous experience in applying **Machine** Vision or **Machine Learning** in robotics or robot application research is a strong merit. new.abb.com
ABB hired **4 people from your company**

Cloud Solution Architect - Data Platform
Microsoft
Kista, Sweden
A Significant depth of technical experience in a number of the following Azure services and technologies: Data Catalog, Data Factory, ...
Your connection **Henrik works at Microsoft**

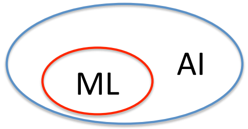
Research Bioinformatician focused on modeling of normal and malignant hematopoiesis...
Karolinska Institutet
Huddinge, SE
In particular we are looking for applicants with math/physics/computer science background and with a desire to learn biology. ki.mynetworkglobal.com
You have **3 connections at Karolinska Institutet**

13

Machine learning vs. AI

Depends on who you ask….

View 1: ML is a subfield of AI



View 2: ML and AI are separate fields

symbolic processing, logic, rule based, expert systems, knowledge representation, databases, planning, reasoning, etc.

learn and generalize from observed data, learn from experience, discover structure in data.

14

What is (machine) learning?

Encyclopaedia Britannica (1964):
"Any relatively permanent change in behaviour resulting from past experience."

Bishop (2006):
"The core objective of a learner is to generalize from its experience."

Wikipedia (2015):
"Machine learning is a scientific discipline that explores the construction and study of algorithms that can learn from data. Such algorithms operate by building a model based on inputs and using that to make predictions or decisions, rather than following only explicitly programmed instructions."

15

How can a machine learn?

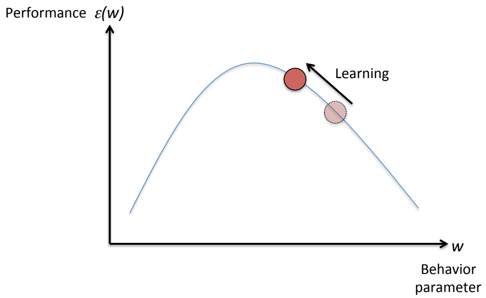
"Any relatively permanent change in behaviour resulting from past experience."

- The "behaviour" of the machine is determined by model parameters.
- "past experience" is previously observed data.

- Machine learning = changing parameter values (behaviour) as a result of observed data (experience).

16

Learning = optimization



Mathematical foundations of machine learning

Optimize parameters for optimal behaviour

Machine Learning

Optimization theory | Probability theory

Multidimensional calculus

Linear algebra

Model uncertainty

Why machine learning?

- Algorithm too complex for a human to implement, but we can easily generate examples of what the algorithm should do.
- Relationships in high-dimensional data too complex for a human to see, but a computer can find these.
- The computer should learn and adapt continuously to new situations.

Big companies are using it

- Microsoft
  http://research.microsoft.com/en-us/about/our-research/machine-learning.aspx
- Google
  http://research.google.com/pubs/ArtificialIntelligenceandMachineLearning.html
- IBM
  http://www.research.ibm.com/cognitive-computing/
- Apple, Facebook, Siemens, Exxon, ... (long list)

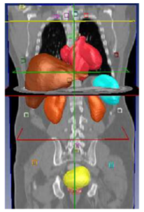
Three main categories of machine learning methods

- **Supervised learning (predictive)**
  Learn to generalize and classify new data based on labelled training data.
  – Pattern recognition
  – Classification
  – Regression
- **Unsupervised learning (descriptive)**
  Discover structure and relationships in complex high-dimensional data.
- **Reinforcement learning (active)**
  Generate policies/strategies that lead to a (possibly delayed) reward. Learning by doing.

Pattern recognition examples

Face recognition

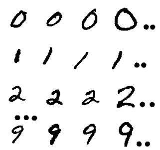
Organ segmentation



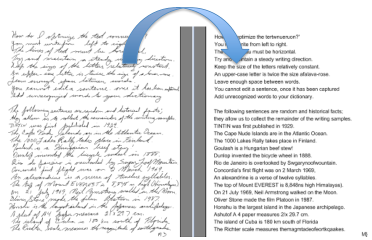
21

Pattern recognition examples

Optical Character Recognition (OCR)



Source: http://blog.damiles.com



22

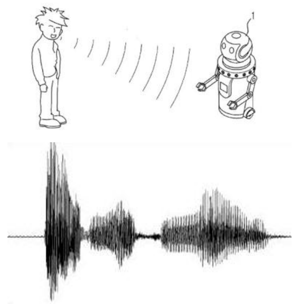
Pattern recognition examples

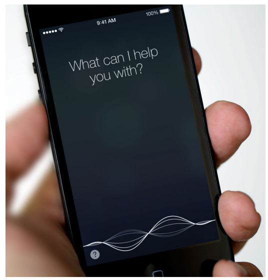
Xbox Kinect – Pose estimation

23

Pattern recognition examples

Speech recognition



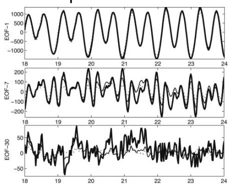


24

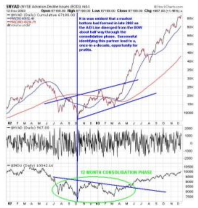
Pattern recognition examples

Prediction and forecasting

Weather and natural phenomena



Financial markets



Pattern recognition examples

Game positions

Spam filters

#1
100% satisfied
4U
Accept credit cards
Act Now!
Additional Income
Affordable
All natural

Movie & music recommendation

Spotify

Features

- A <u>feature</u> is a measurement or scalar number that describes some aspect of a phenomenon or object
  - Size, length, shape, velocity
  - Intensity and color (RGB)
  - Position (x,y)
  - Signal frequency
  - Sensor measurements (e.g., temperature)
  - Game piece present at certain location (yes/no)
  - Word present in an email (yes/no)
- <u>Feature extraction</u> is the process of measuring features from data.

Features – Iris dataset

Iris setosa

Iris versicolor

Iris virginica

Fisher's Iris Data

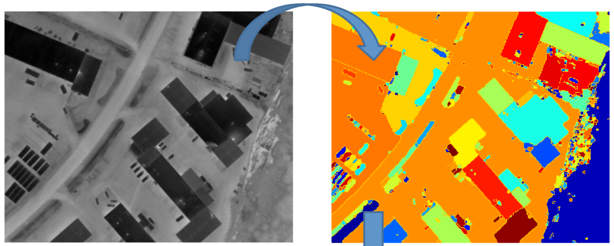
Sepal Length	Sepal Width	Petal Length	Petal Width	Species
5.7	4.4	1.5	0.4	*I. setosa*
5.8	2.6	4.0	1.2	*I. versicolor*
5.8	2.7	5.1	1.9	*I. virginica*

From Wikipedia

Feature vectors: $\mathbf{x}_1 = \begin{bmatrix} 5.7 \\ 4.4 \\ 1.5 \\ 0.4 \end{bmatrix}$ $\mathbf{x}_2 = \begin{bmatrix} 5.8 \\ 2.6 \\ 4.0 \\ 1.2 \end{bmatrix}$ $\mathbf{x}_3 = \begin{bmatrix} 5.8 \\ 2.7 \\ 5.1 \\ 1.9 \end{bmatrix}$

Features – Image classification

Watershed transform

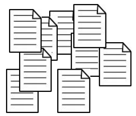


Extract intensity & shape features

$$\mathbf{x} = \begin{bmatrix} \text{Mean intensity} \\ \text{Std intensity} \\ \text{Border regularity} \\ \text{Compactness} \end{bmatrix}$$

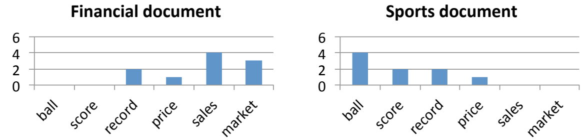
29

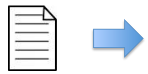
Features – Document analysis



Articles, mail, web pages, ...

Dictionary:
{'ball', 'score', 'record', 'price', 'sales', 'market'}



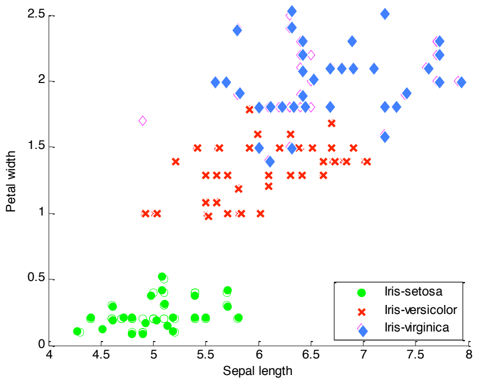


$$\mathbf{x} = \begin{bmatrix} \#'ball' \\ \#'score' \\ \#'record' \\ \#'price' \\ \#'sales' \\ \#'market' \end{bmatrix}$$

Bag-of-words model

30

Feature space



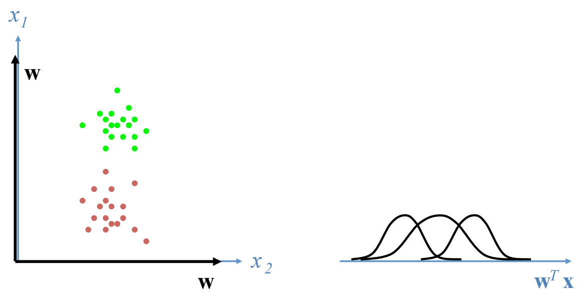
Iris setosa

Iris versicolor

Iris virginica

31

Representation



32

Supervised learning

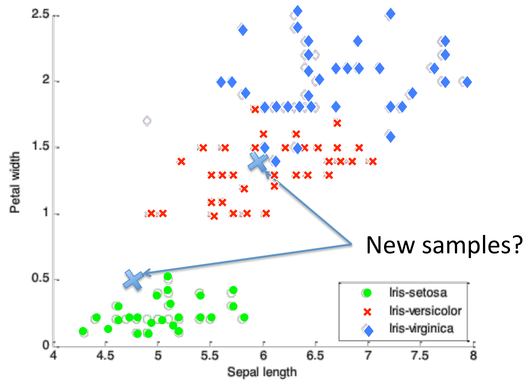
- **Task:** Learn to predict/classify new data from labeled examples.
- **Input:** Training data examples $\{\mathbf{x}_i, y_i\}$ i=1...N, where $\mathbf{x}_i$ is a feature vector and y_i is a class label in the set Ω.
- **Output:** A function $f(\mathbf{x};w_1,\ldots,w_k) \rightarrow \Omega$

Find a function *f* and adjust the parameters $w_1,\ldots,w_k$ so that new feature vectors are classified correctly. ***Generalization!!***

Classification vs. regression vs. ranking

- **Classification**: Select one of a discrete set of classes (the set Ω is discrete).
  - Which horse is going to win this race?
  - Which letter does this image depict?
  - Is this email spam (yes/no)?
- **Regression**: Learn to predict a continuous value ($\Omega = \mathbb{R}$).
  - Learn to predict the temperature tomorrow.
  - What is the *probability* that this image depicts the letter 'a'?
- **Ranking**: Learn to rank a set of items ($\Omega = \mathbb{R}$).
  - Rank webpages, movies, etc.

Example



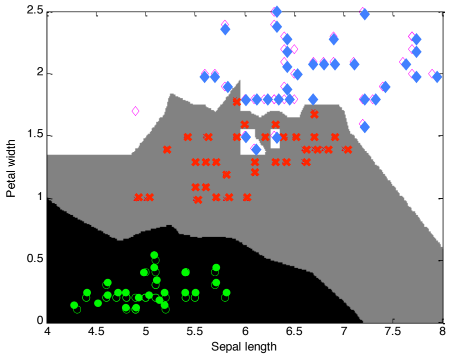
k-Nearest Neighbours (k-NN)

- Save all training data.
- For a new case, find similar examples among the training data.
- Requires a similarity measure (metric), for example the Euclidian distance

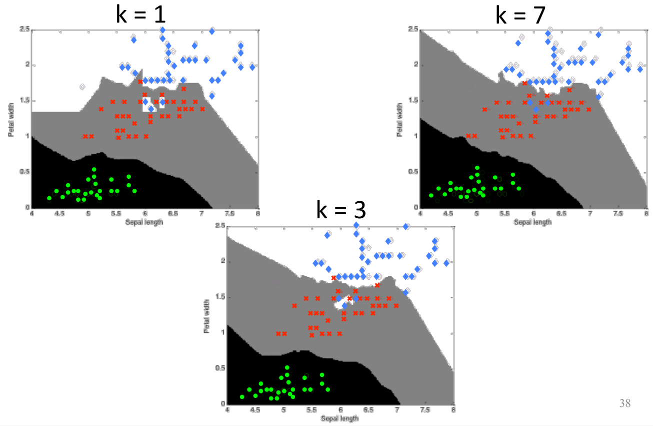
$$\|\mathbf{x}-\mathbf{y}\| = \sqrt{\sum_i (x_i - y_i)^2}$$

- A majority vote among the *k* nearest neighbours decides the class, where *k* can be 1,2,3,4...

Classification areas
k = 1



Classification areas



Pros and cons of k-NN

- Very simple – no “training“ or modeling required
- Must store all training data – problem for large data sets:

$$f(\mathbf{x}; \underbrace{w_1,\dots,w_k}) \rightarrow \Omega$$

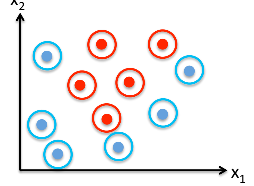
Parameters equal to training data $\mathbf{x}_i$

- Slow classification for large data sets – must compare new samples with all stored samples.

Generalization!

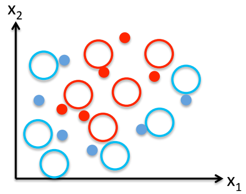
“The core objective of a learner is to generalize from its experience.”

Classifying training data is trivial. Memorization!



$f(\mathbf{x}; w_1,\dots,w_k) \rightarrow \Omega$

Classifying **new data** is the challenge!



Generalization error!

Evaluating classifiers

- How can we compare the performance of different classifiers?
- What happens if we use the same data for training and evaluation?
- How can we train and test a classifier if we only have a finite amount of collected data?

Confusion matrix


Iris setosa


Iris versicolor


Iris virginica

Predicted class

Actual class	Setosa	Versicol.	Virginica
Setosa	50	0	0
Versicol.	0	45	5
Virginica	0	7	43

Accuracy: $\frac{50+45+43}{150} = 92\%$

Training data vs. test data

- A classifier must be able to <u>generalize</u>, i.e., it must be tested using <u>previously unseen</u> data.
- Evaluating using training data will give an overly optimistic accuracy.
- Three ways to perform the evaluation:
  - Hold out
  - Cross validation
  - Leave one out

Hold out

- Simplest approach, hold out one part of the entire data set as test data.

Training data	Test data

n-fold Cross-Validation

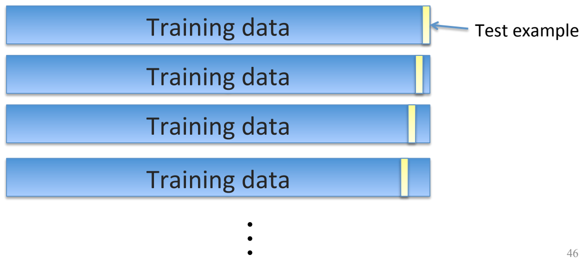
- Divide data set into n segments. Train using n-1 segments and evaluate using the n:th.
- Example of 3-fold Cross-Validation:

Training data	Training data	Test data
Training data	Test data	Training data
Test data	Training data	Training data

45

Leave-one-out

- Extreme case of Cross-Validation: Use all data but one example for training and use the last one to evaluate



46