



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

Μηχανική μάθηση

Ενότητα 1: Introduction to machine learning

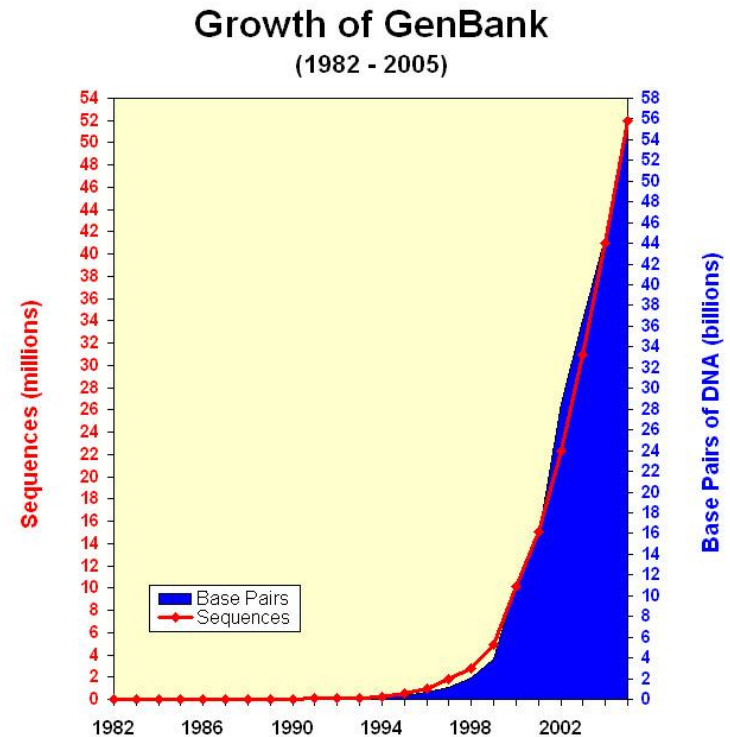
Ιωάννης Τσαμαρδίνος
Τμήμα Επιστήμης Υπολογιστών

Machine Learning

- The study of algorithms and systems that improve their performance with experience (Mitchell)
- Experience?
- Experience = data / measurements / observations

Data Everywhere

- Corporate Databases
 - ▣ Banks, insurance companies
- Web
 - ▣ Spam, facebook, google
- Scientific databases
 - ▣ GenBank
- Own Data



Source : GenBank website

Where to Use Machine Learning

- You have past data, you want to predict the future
- You have data, you want to make sense out of them (find useful patterns)
- You have a problem it's hard to find an algorithm for
 - ▣ Gather some input-output pairs, learn the mapping
- Measurements + intelligent behavior usually lead to some form of Machine Learning

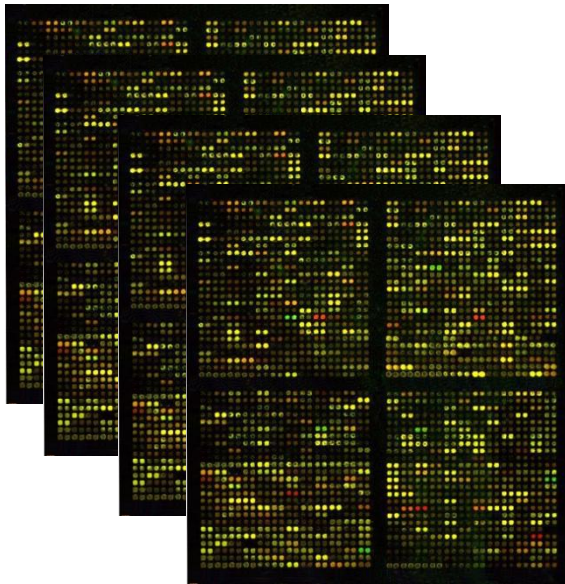
Machine Learning as a Field

- Hot, hot, hot!
- ML techniques and algorithms applied in
 - ▣ Bioinformatics, medical informatics, computational biology
 - ▣ Mining the Web
 - ▣ Speech recognition
 - ▣ Natural language processing
 - ▣ Computer vision
 - ▣ Robot control
 - ▣ ...

Machine Learning as a Field

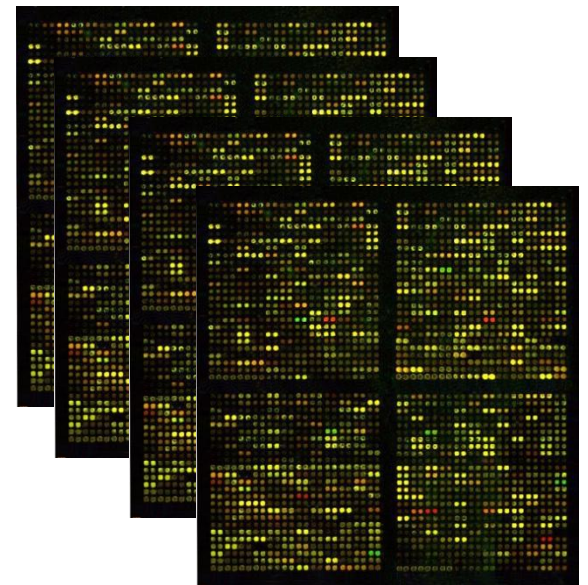
- Relative good: Jobs, tasks, funding
- Very active field
- Significant breakthroughs last 2 decades

Clinical Bioinformatics

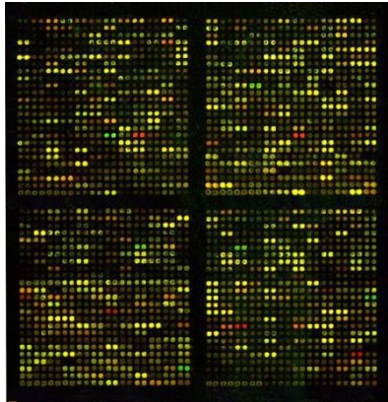


Control group of
normal tissue

Cases of
cancerous tissue

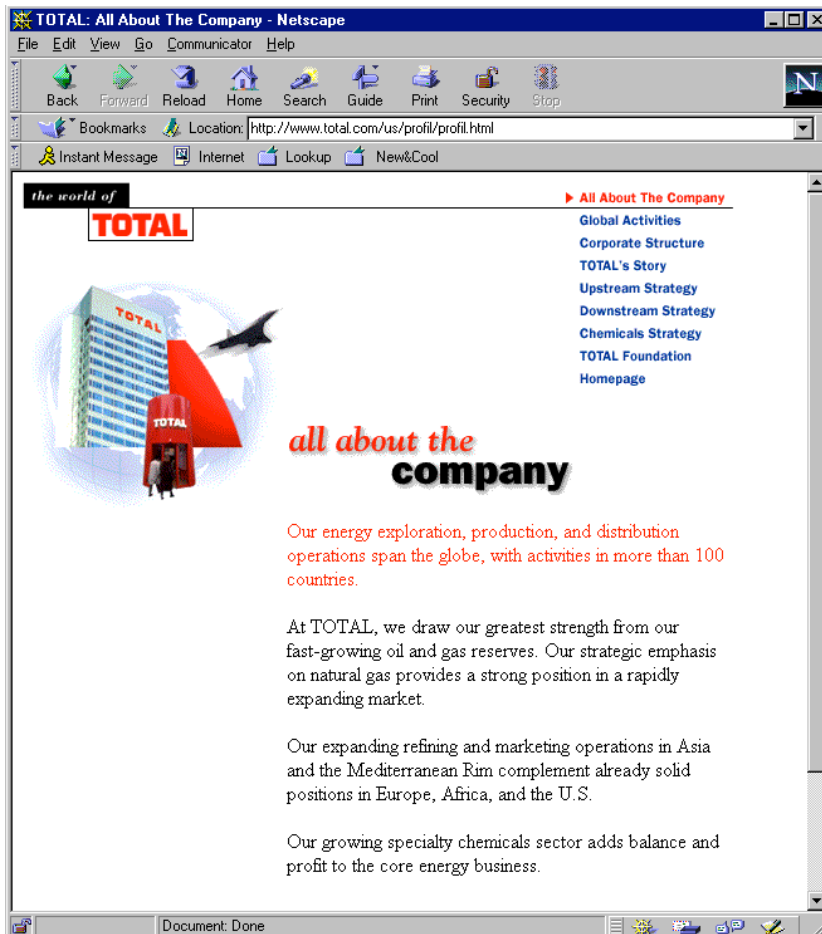


Clinical Bioinformatics



Is this new tissue (patient)
from an early cancerous
tissue? (prediction /
diagnosis)

Text Classification



Is this a web page of

...

a company?

Personal home page?

Academic institution?

...

Spam Detection

debt consalidation
benton delnaz [55dilemma3000@mail2spectrum.com]
To: tsamard@cs.pitt.edu
Cc:

Its time to completely eradicate your debts!

ELIMINATE your carddebts and ALL other unsecured debts

Visit www.bestfirstcredit.com

- * NO more payments to creditors
- * YOUR long term credit will NOT be affected.
- * NO confrontation's

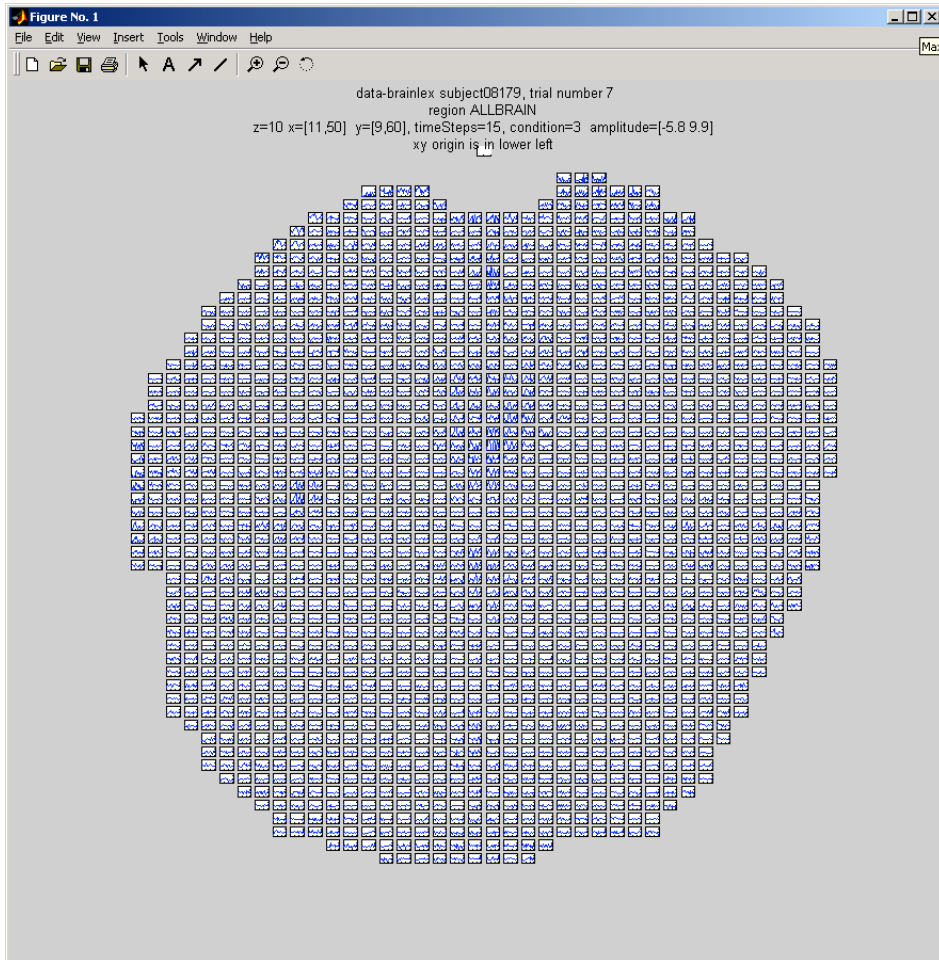
Visit www.bestfirstcredit.com

- * 10K minimum combined debts required for eligibility.
- * US residents only.

Is this a spam email?

You bet it is!

Mind Reading



Is this person thinking
of a verb or a
noun?

[Rustandi et al., 2005]

Autonomous Systems



Source : www.in.gr

Ένα μικρό ελικόπτερο εξοπλισμένο με σύστημα τεχνητής νοημοσύνη μαθαίνει να πετά μόνο του στην πανεπιστημιούπολη του Στάνφορντ με την προοπτική να χρησιμοποιηθεί από το στρατό, από την αστυνομία ή από τις δασικές υπηρεσίες.

«Το ελικόπτερο αυτό μπορεί να κάνει μανούβρες στο όριο των δυνατοτήτων των ελικοπτέρων» υπερηφανεύτηκε ο διδακτορικός φοιτητής Ανταμ Κόοατς, μέλος της ερευνητικής ομάδας.

Όπως αναφέρει την Τρίτη το Reuters, το αυτόνομο ελικόπτερο κατευθύνεται από έναν υπολογιστή στο έδαφος ο οποίος μαθαίνει να πιλοτάρει παρακολουθώντας τις πτήσεις τηλεκατευθυνόμενων ελικοπτέρων μέσω δύο επίγειων καμερών

Το ίδιο το αεροσκάφος, κόστους περίπου 4.000 δολαρίων έκαστο, διαθέτει επιταχυνσιόμετρο, γυροσκόπιο και μαγνητόμετρο για να αντιλαμβάνεται τον προσανατολισμό του, καθώς και δέκτη GPS.

Ενδιαφέρον για τη νέα τεχνολογία έχουν εκδηλώσει εταιρείες που δραστηριοποιούνται σε εφαρμογές επιτήρησης και χαρτογράφησης, καθώς και επιστήμονες που σκοπεύουν να το χρησιμοποιήσουν για την παρακολούθηση πυρκαγιών σε δύσβατα δάση.

Types of Learning Paradigms

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Supervised Learning

- Learn from examples
- Would like to be able to predict an outcome of interest y for an object x
- Learn function $y = f(x)$
- For example, x is a patient, y is an indicator of a disease (0 if the patient does not have the disease and 1 otherwise)
- We are given data with pairs $\{ \langle x_i, y_i \rangle : i=1, \dots, n \}$,
 - x_i the representation of an object
 - y_i the representation of a known outcome
- Learn the function $y = f(x)$ that generalizes from the data the “best” (has minimum average error)

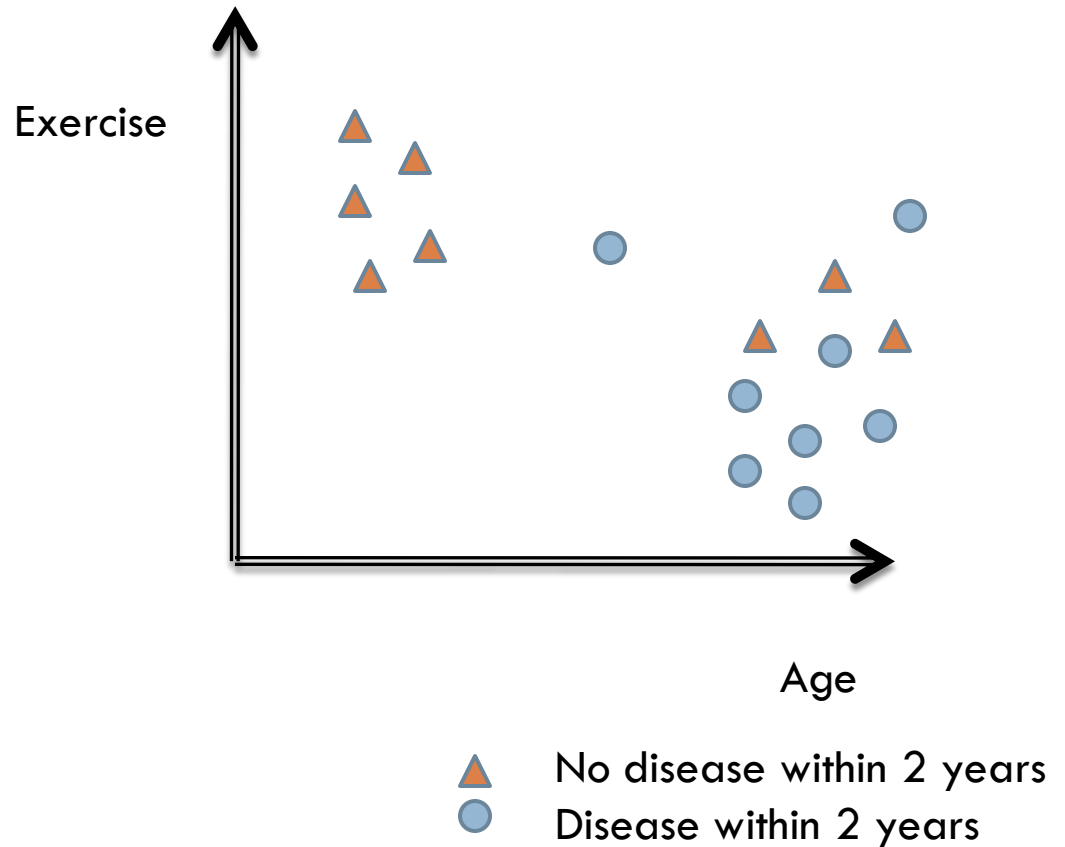
Supervised Learning: Example

- 2-year prognosis of Parkinson's disease
- Representation of a patient
 - Measure the age and times per week she exercises
 - Vector of discrete and continuous variables
- Representation of output: Yes, No
- Generalize the examples to learn to provide a 2-year prognosis with the minimum possible error

Example	X^1 (Age)	X^2 (Exercises) (times per week)	Y (Parkinson's within 2 years)
1	65	0	Yes
2	58	2	No
...			
N	60	5	No

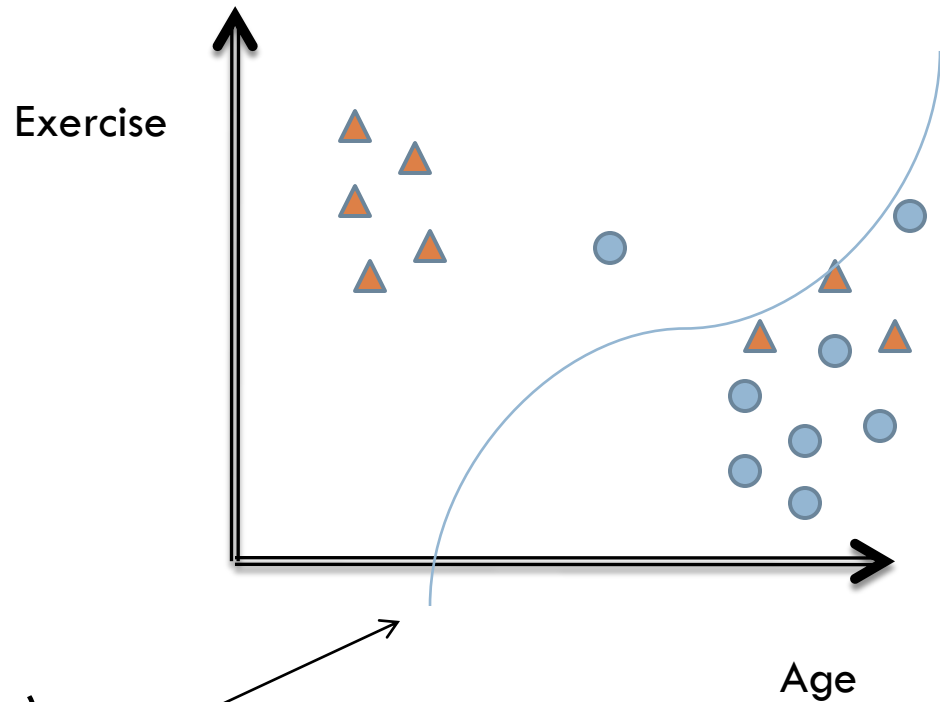
Supervised Learning

How to generalize the data and provide prognosis for all (new) patients?



Supervised Learning

How to generalize the data and provide prognosis for all (new) patients?



A possible function $y = f(x)$:
Everything to the left takes value No, and to the right Yes

- ▲ No disease within 2 years
- Disease within 2 years

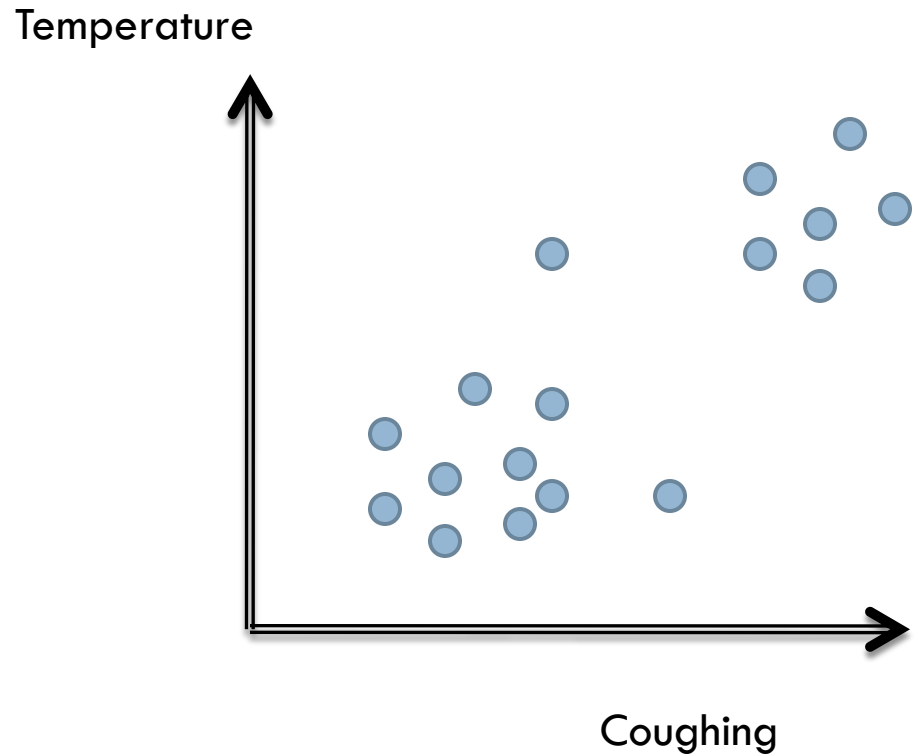
Unsupervised Learning

- Learn interesting patterns, structure and groups of objects
- Depends on the definition of “interesting”

Example	X^1 (Temperature)	X^2 (Amount of Coughing per day)
1	35,6	15
2	40,1	20
...	36,6	1
N	36,6	1

Unsupervised Learning

- Find best grouping
- There seem to be two groups
- Hmm
- Name one “people with flu” and the other ... “normals”
- Now, let’s find out why some people get the flu



Reinforcement Learning

- Learn how to act in different situations
- Again, learns from examples, but examples of action and behavior
- Data: triplets of O, A, R
 - ▣ Represent my historic observations and actions O, A
 - ▣ Represent my historic “reward”, R
- Generalize from the data to learn the function
 - ▣ $y = f(O, A, R)$, y the next action to take

Syllabus

- Wide range of Machine Learning techniques, both basic and state-of-the-art
 - ▣ Decision Trees, Naïve Bayes, K-Nearest Neighbors, Support Vector Machines, Artificial Neural Networks
- Evaluation of performance and model selection
 - ▣ Cross-validation, nested cross-validation, bootstrapping, Receiving Operating Curves, accuracy
- Application methodology and practice
- Feature Selection
- Clustering
- Causal Discovery!

Prerequisites

- Probabilities
 - ▣ Will somewhat be reviewed in brief
- Algorithms
 - ▣ Will not be reviewed
- Optimization Theory
 - ▣ Brief introduction and review
- Programming
 - ▣ in Matlab mostly
 - ▣ Check you can use Matlab at UOC facilities

Recitations

- Necessary! and useful
 - Review material
 - Present background
 - Answer questions
 - Do practical assignments
- First few recitations: probabilities + Matlab

Supervised Learning

- The problem of induction
 - ▣ Induction: from a few examples, infer properties of all objects
 - All observed crows are black
 - Therefore: all crows are black
- Hume's argument about induction
- Induction is the basis of all natural sciences
- Pictorial examples of the problem of induction

What to do about Induction

- Turn induction into deduction by adding assumptions
 - The physicist:
 - The physical laws now will be the same in the future
 - The physical laws here, are the same as everywhere
 - The mathematician/computer scientist
 - The function I am trying to learn is smooth: small changes in the input, usually have small changes in the output
 - The statistician:
 - My distribution is normal

General Principles of Learning

- Ockham's razor: "Causes shall not be multiplied beyond necessity"
- Among all hypotheses that explain the data the same, choose the simplest
- Definition of simplicity?

General Principles of Learning

□ Minimum Description Length

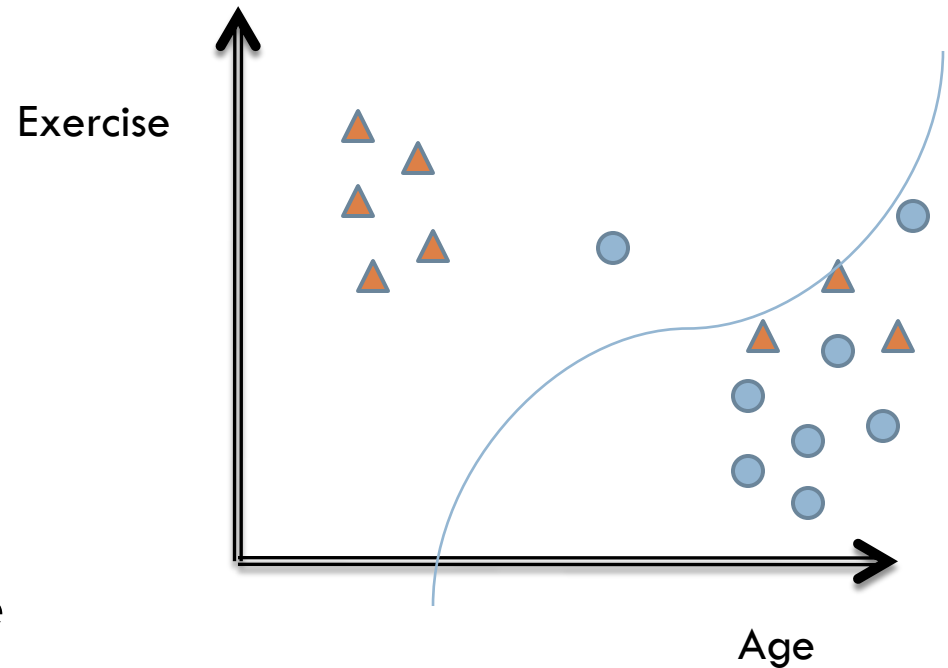
- Formalizes Ockham's Razor principle and generalizes it to compare hypothesis/models that explain the data approximately the same
- The hypothesis/model requires some memory to represent M (given a finite alphabet of symbols)
- The errors / exceptions induced by the model require some memory to represent D
- Choose the hypothesis that minimizes the communication bits of $M+D$

General Principles of Learning

- Regularization (“make it regular”)
 - Formalizes Ockham’s Razor principle
 - Introduces penalties to complexity
 - Fit the data with a penalty for increased complexity

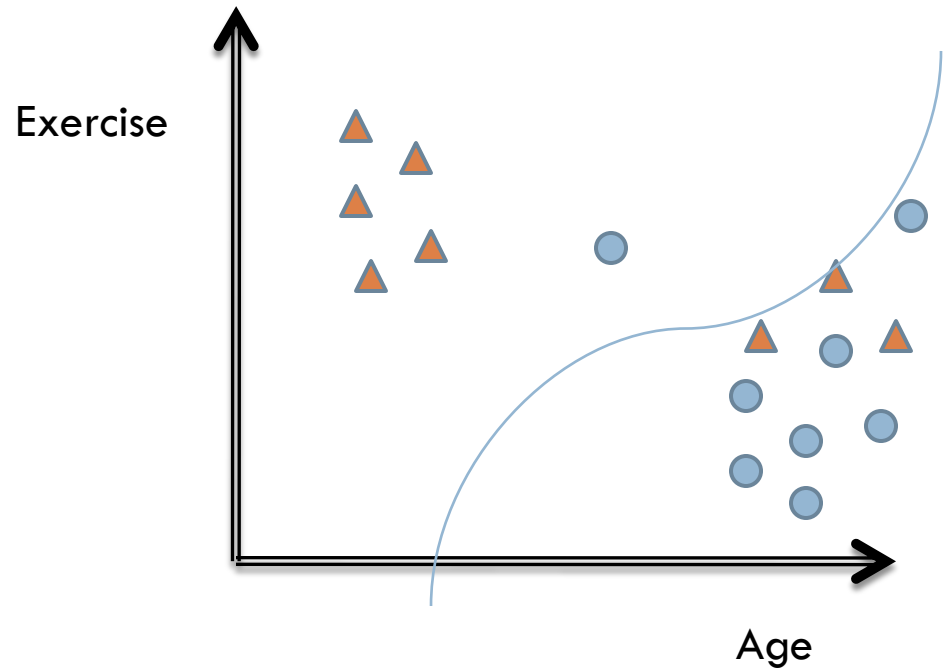
Model Space

- Want to learn a function $y=f(x)$
- Each such function is called hypothesis or model
- Hypothesis space or model space
 - ▣ Space of all possible models that we are considering



Inductive Bias

- The set of assumptions that turns the inductive problem into a deductive one
 - Model Space induces a bias when it excludes functions
 - Preferences over models (e.g. using Minimum Description Length) induce a bias
 - Search method for the best model induces a bias



You cannot learn without inductive bias! (No Free Lunch Theorem)

Classification versus Regression

- $y = f(x)$
- Classification
 - y 's domain is finite
 - Example: $y \in \{\text{Yes, No}\}$
- Regression
 - y 's domain is infinite (typically continuous)
 - Example: $y \in \mathbb{R}$
- In this class, we mostly deal with classification

Things to Know

- Readings Chapter 1, Machine Learning book
- Types of different learning paradigms
 - ▣ How they are set up mathematically
 - ▣ Examples from each
- What's the problem of induction
- What are some general principles of learning
- Inductive Bias + Model Space
- Classification versus Regression

Τέλος Ενότητας



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Κρήτης**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.





Σημειώματα

Σημείωμα αδειοδότησης (1)

- Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Μη Εμπορική Χρήση, Όχι Παράγωγο Έργο 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



[1] <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Σημείωμα αδειοδότησης (2)

- Ως **Μη Εμπορική** ορίζεται η χρήση:
 - που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
 - που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
 - που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο
- Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.

Σημείωμα Αναφοράς

Copyright Πανεπιστήμιο Κρήτης, Ιωάννης Τσαμαρδίνος 2015.
«Μηχανική Μάθηση. Introduction to machine learning».
Έκδοση: 1.0. Ηράκλειο 2015. Διαθέσιμο από τη δικτυακή
διεύθυνση:
<https://opencourses.uoc.gr/courses/course/view.php?id=362>.

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.