Machine Learning

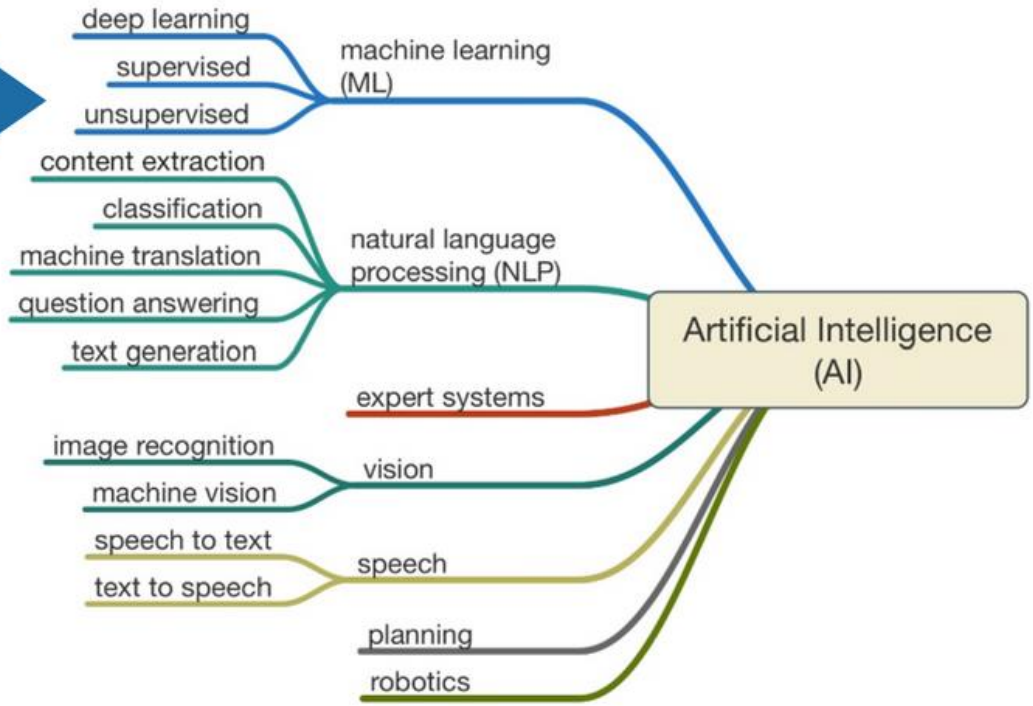
Introduction

by *Marina Barsky*

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Machine learning teaches machines to learn to carry out tasks by themselves, without given explicit instructions

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements



- Machine learning is a subfield of artificial intelligence (AI) concerned with algorithms that allow computers to *learn*
- An algorithm is given a set of data and if data is non-random, it contains patterns
- Based on these patterns, ML algorithm builds a generalized model of data
- That allows it to make predictions about other data that it might see in the future

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

This is a lot of experience to learn from!

World Data Centre for Climate (WDCC)

220 terabytes of data on climate research and climatic trends,

110 terabytes worth of climate simulation data.

6 petabytes worth of additional information stored on tapes.

AT&T

323 terabytes of information

1.9 trillion phone call records

Google

91 million searches per day, After a year more than 33 trillion database entries.

ML algorithms incorporate mental models of learning

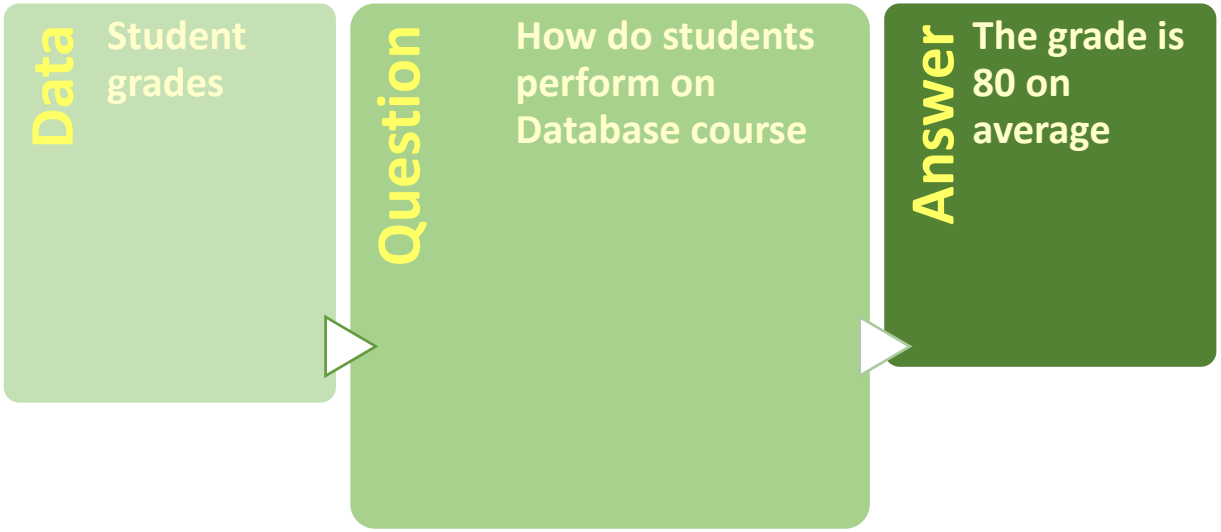
- What do we mean by learning?
  - Based on the previous experiences assign a label to a new object
  - Group similar things together into a single category
  - Identify repeating patterns

The ML algorithms use similar ideas:

- The previous experiences are encoded as a set of data points
- The inference of a label or a category is automatic

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

What is (not) Machine Learning

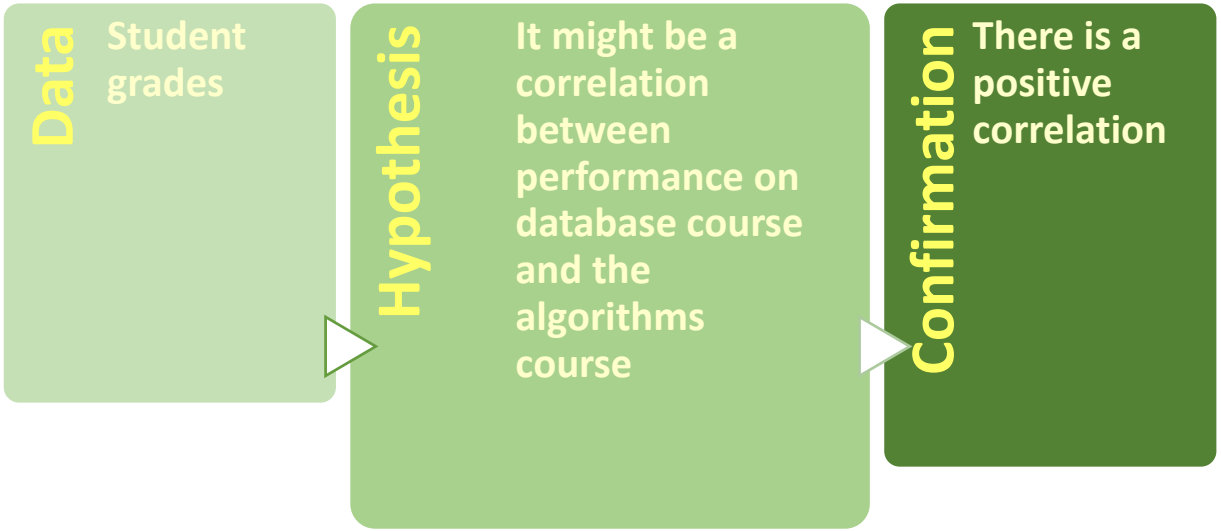


Not ML

- data manipulation (query)

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

What is (not) Machine Learning

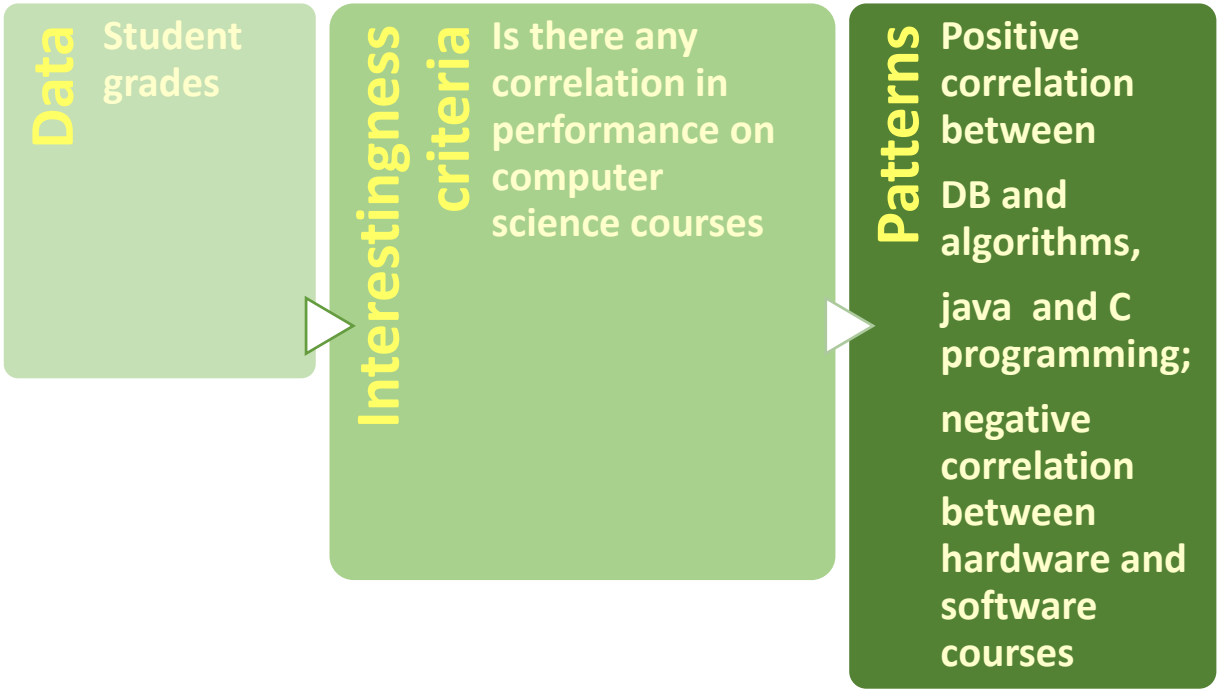


Not ML problem

- statistics (hypothesis testing)

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

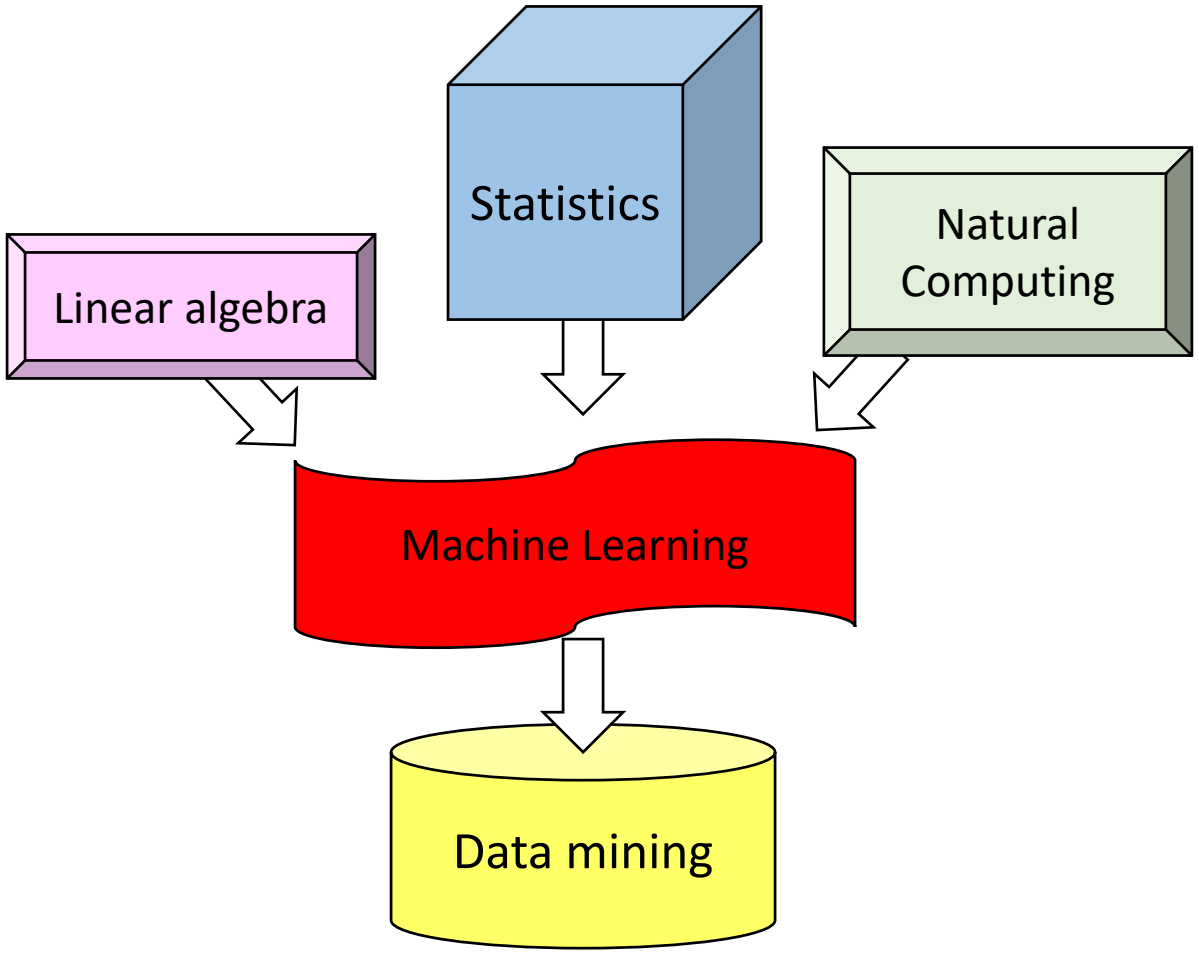
What is (not) Machine Learning



Machine learning!

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Not a separate discipline



Combines ideas from all these disciplines into practical ***algorithms***

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Facial recognition?
http://www.pictriev.com/

This course:
Computer Science part of Machine Learning

- We focus on ***algorithms***
- By the end you understand the ideas behind ML algorithms: both mental and mathematical
- You will experiment with programs based on these algorithms and see by yourselves whether machines can or cannot learn

Far-reaching Course Objectives

- Develop interest in math as a tool for learning about the world
- Learn how to handle ambiguity
- Formalize mental models of learning
- Incorporate ML algorithms in your smart web applications
- Invent new ML approaches and new algorithms

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Why study Machine Learning? 1/3

Get **competitive advantage** in business

- Google uses web links to rank pages, it gathers your every click and learns to adapt its search to your preferences
- Amazon and Netflix use information about the things people buy or watch to learn which people or items are similar to one another, and then make recommendations
- Pandora and Last.fm use your ratings of songs to create custom radio stations with music they think you will enjoy
- The predictions made by the Hollywood Stock Exchange are routinely better than those made by individual experts
- eHarmony uses information collected from participants to determine who would be a good match

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Why study Machine Learning? 2/3

In **science**:

- Classify faint galaxies
- Find similar gene expressions for different drug treatments
- Predict structure of a chemical from magnetic resonance data

…

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Why study Machine Learning? 3/3

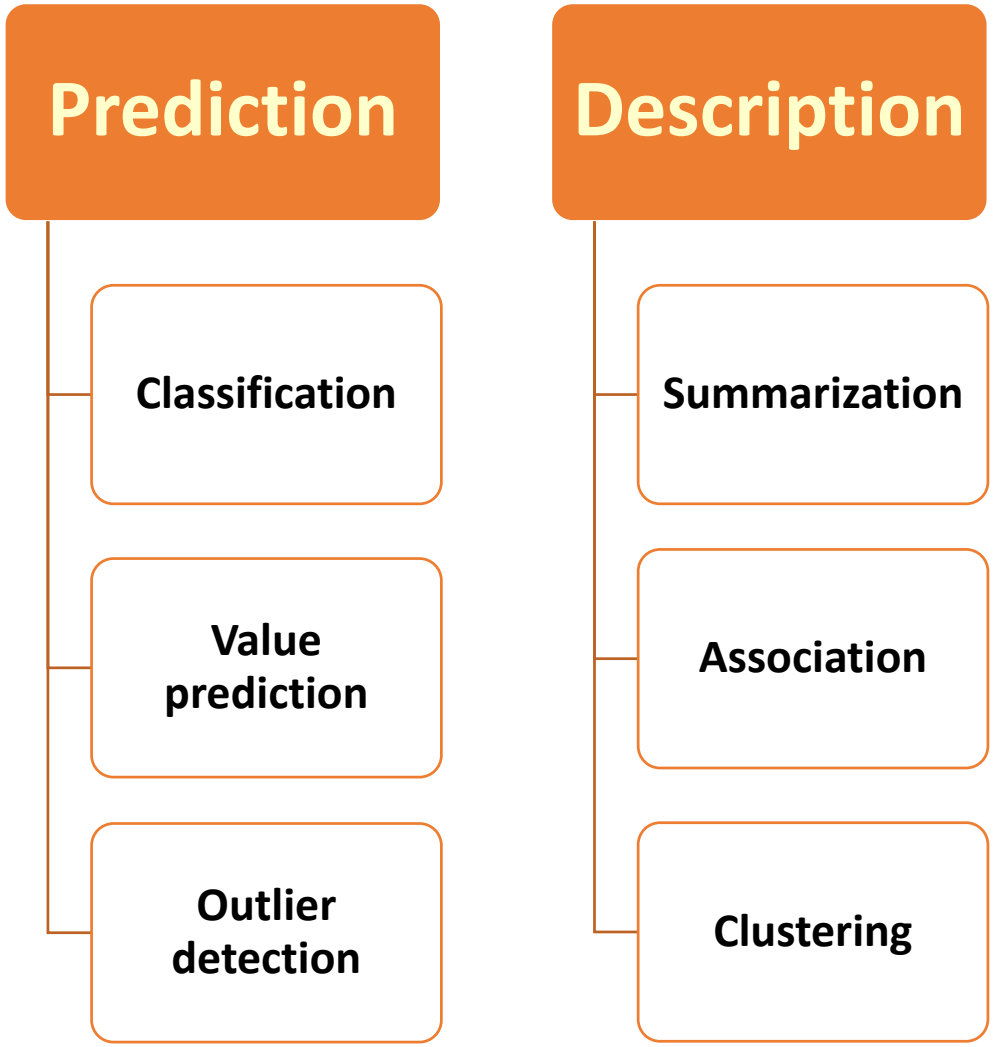
Automate everyday tasks

- Show to the algorithm which messages you consider a spam, and the task of separating spam can be carried out automatically
- Collect only positive or only negative news articles
- …

Once you learn about a few machine-learning algorithms, you'll start seeing places to apply them just about everywhere

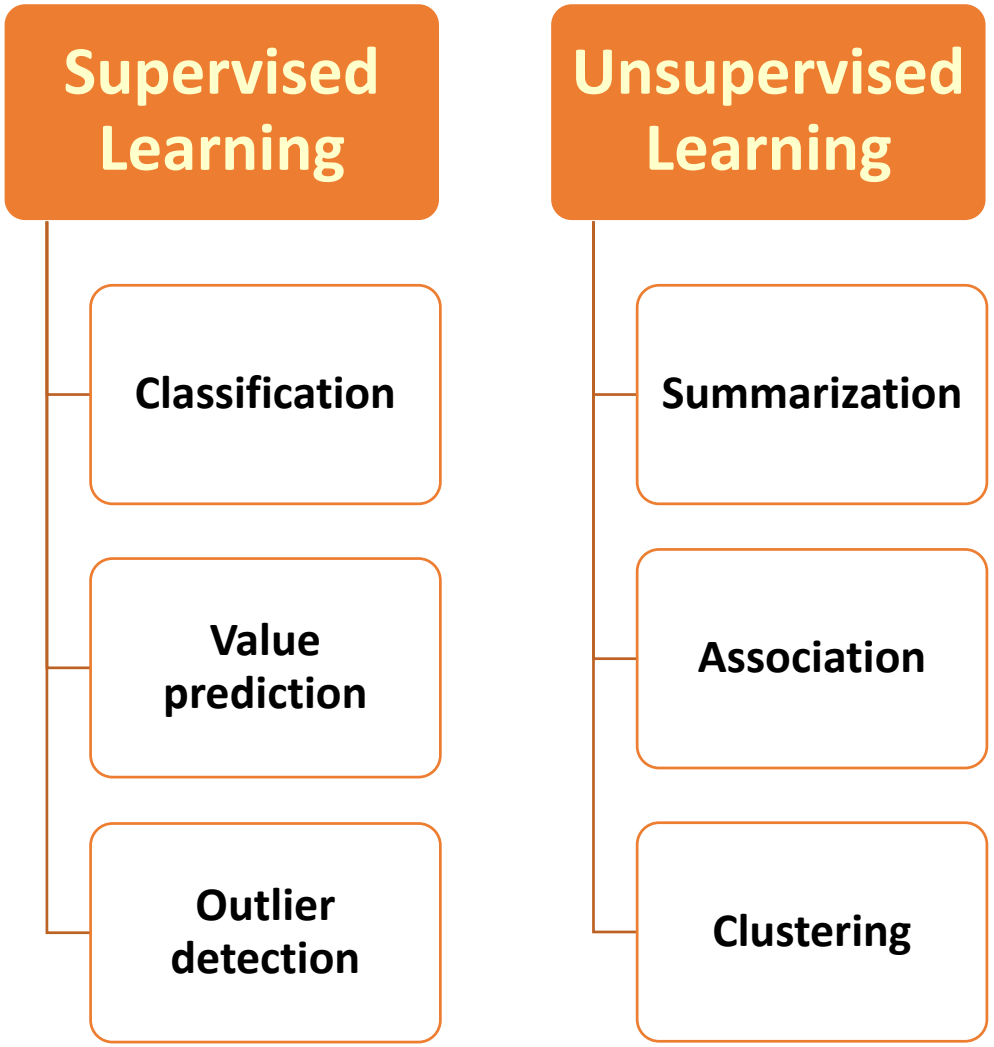
- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Task types



- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Task types



- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Task of type 1: Classification

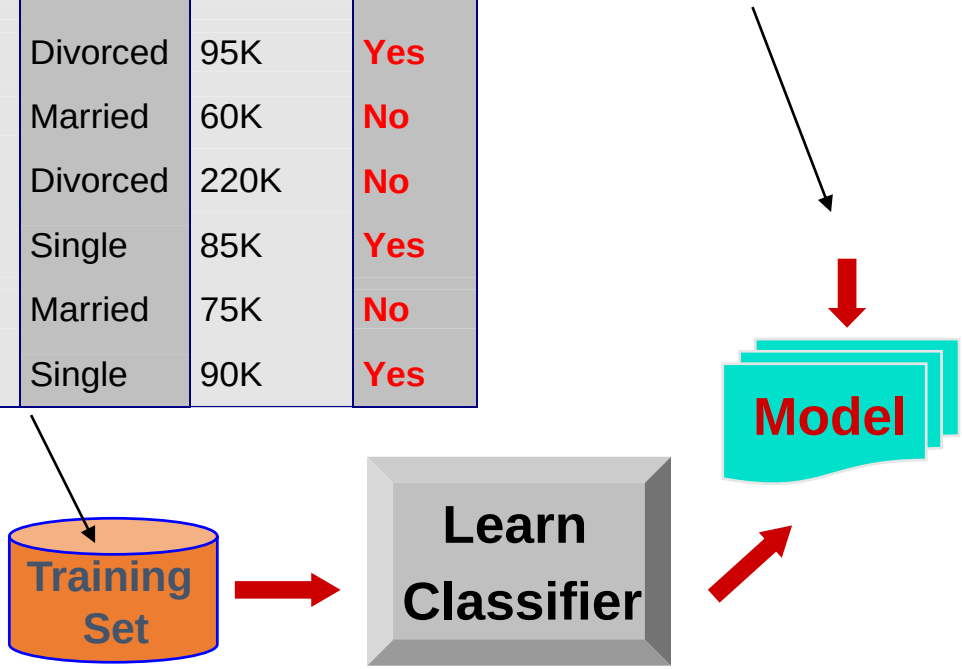
- Given a collection of records (*training set*)
  - Each record contains a set of *attributes*, one of the attributes is the *class*.
- Find ("learn") a *model* for the class attribute as a function of the values of the other attributes.
- Goal: **previously unseen** records should be assigned a class as accurately as possible.

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Classification example

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Refund	Marital Status	Taxable Income	Cheat
No	Single	75K	?
Yes	Married	50K	?
No	Married	150K	?
Yes	Divorced	90K	?
No	Single	40K	?
No	Married	80K	?



- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Solving classification problem

My neighbour dataset

class label

Temp	Precip	Day	Shop	Clothes	
25	None	Sat	No	Casual	**Walk**
-5	Snow	Mon	Yes	Casual	**Drive**
15	Snow	Mon	Yes	Casual	**Walk**

(Adopted from Leslie Kaelbling's example in the MIT courseware)

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Classification problem

class label

Temp	Precip	Day	Shop	Clothes	
25	None	Sat	No	Casual	**Walk**
-5	Snow	Mon	Yes	Casual	**Drive**
15	Snow	Mon	Yes	Casual	**Walk**
-5	Snow	Mon	Yes	Casual	**?**

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Classification problem: memory

class label

Temp	Precip	Day	Shop	Clothes	
25	None	Sat	No	Casual	**Walk**
-5	Snow	Mon	Yes	Casual	**Drive**
15	Snow	Mon	Yes	Casual	**Walk**
-5	Snow	Mon	Yes	Casual	**Drive**

(Adapted from Leslie Kaelbling's example in the MIT courseware)

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Classification problem: noise

Temp	**Precip**	**Day**	**Clothes**	
25	None	Sat	Casual	**Walk**
25	None	Sat	Casual	**Walk**
25	None	Sat	Casual	**Drive**
25	None	Sat	Casual	**Drive**
25	None	Sat	Casual	**Walk**
25	None	Sat	Casual	**Walk**
25	None	Sat	Casual	**Walk**
25	None	Sat	Casual	**?**

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Classification problem: averaging

Temp	Precip	Day	Clothes	
25	None	Sat	Casual	**Walk**
25	None	Sat	Casual	**Walk**
25	None	Sat	Casual	**Drive**
25	None	Sat	Casual	**Drive**
25	None	Sat	Casual	**Walk**
25	None	Sat	Casual	**Walk**
25	None	Sat	Casual	**Walk**
25	None	Sat	Casual	**Walk**

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Classification problem: generalization

Temp	**Precip**	**Day**	**Clothes**	
22	None	Fri	Casual	**Walk**
3	None	Sun	Casual	**Walk**
10	Rain	Wed	Casual	**Walk**
30	None	Mon	Casual	**Drive**
20	None	Sat	Formal	**Drive**
25	None	Sat	Casual	**Drive**
-5	Snow	Mon	Casual	**Drive**
27	None	Tue	Casual	**Drive**
24	Rain	Mon	Casual	**?**

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Learning to predict class label

Three different problems involved in learning:

- memory
- averaging
- generalization.

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Type 2. Explorations

Tid	Refund	Marital Status	Taxable Income
1	Yes	Single	125K
2	No	Married	100K
3	No	Single	70K
4	Yes	Married	120K
5	No	Divorced	95K
6	No	Married	60K
7	Yes	Divorced	220K
8	No	Single	85K
9	No	Married	75K
10	No	Single	90K

Discover groups, no class labels

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Task of type 2. Associations

The Market-Basket Model

- A large set of *items*, e.g., things sold in a supermarket.
- A large set of *baskets*, each of which is a small set of the items, e.g., the things one customer buys in one transaction.

Fundamental problem

- What sets of items are often bought together?

Application

- If a large number of baskets contain both hot dogs and mustard, we can use this information. How?

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Solving association problem: market basket

	Transactions
1	{bread, milk, peanut butter}
2	{bread, milk}
3	{beer, potato chips}
4	{beer, diapers}
5	{beer, milk, diapers}
6	{bread, milk, yogurt}
7	{beer, bread, diapers}
8	{bread, milk, jelly}
9	{beer, cigarettes, diapers}
10	{bread, milk}

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Association problem

	Transactions
1	{**bread**, **milk**, peanut butter}
2	{**bread**, **milk**}
3	{**beer**, potato chips}
4	{**beer**, **diapers**}
5	{**beer**, **milk**, **diapers**}
6	{**bread**, **milk**, yogurt}
7	{**beer**, **bread**, **diapers**}
8	{**bread**, **milk**, jelly}
9	{**beer**, cigarettes, **diapers**}
10	{**bread**, **milk**}

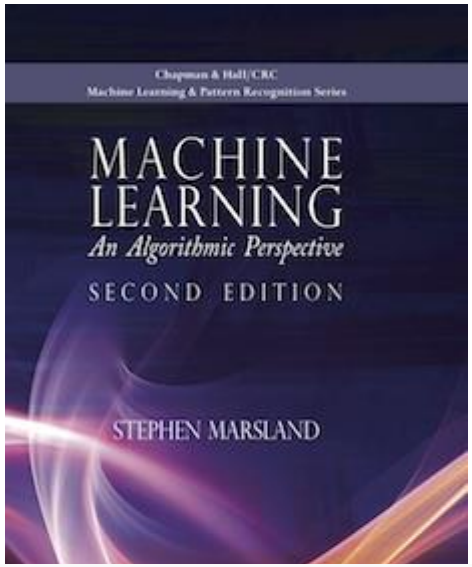
- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Beer and diapers?

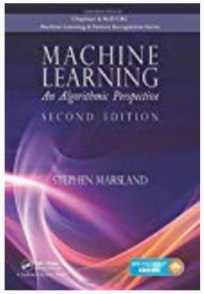
	Transactions
1	{bread, milk, peanut butter}
2	{bread, milk}
3	{**beer**, potato chips}
4	{**beer**, **diapers**}
5	{**beer**, milk, **diapers**}
6	{bread, milk, yogurt}
7	{**beer**, bread, **diapers**}
8	{bread, milk, jelly}
9	{**beer**, cigarettes, **diapers**}
10	{bread, milk}

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Amazon example



Customers who viewed **Machine Learning: An Algorithmic Perspective (Chapman & Hall/Crc Machine...** also viewed



Machine Learning: An Algorithmic Perspective, Second Edition

★★★★☆ 46

$69.29

prime

48 used and new from $59.61

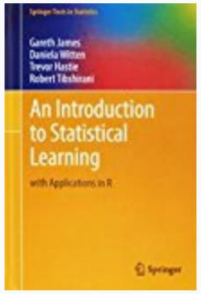
Hands-On Machine Learning with Scikit-Learn and TensorFlow:

★★★★½ 251

$29.35

prime

85 used and new from $22.86



An Introduction to Statistical Learning: with Applications in R (Springer Texts

★★★★½ 200

$49.60

prime

20 used and new from $39.95

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Amazon example ?

Customers Who Bought This Item Also Bought

Revere Polished Aluminum 8-Inch Nonstick Skillet by Revere

★★★★½ (16)

$14.99

Pyrex Smart Essentials 8-Piece Mixing Bowl Set by Pyrex

★★★★½ (66)

$26.82

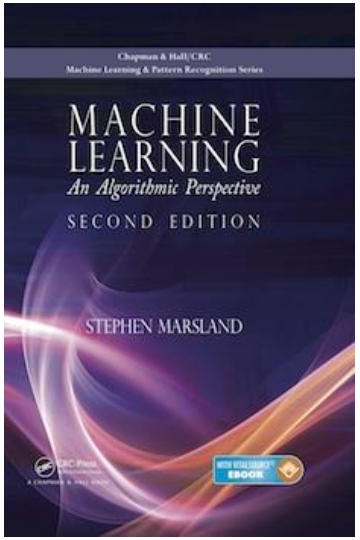
Kodak Portra 400 Professional ISO 400, 35mm, 36 Exposures, Color...

★★★★½ (5)

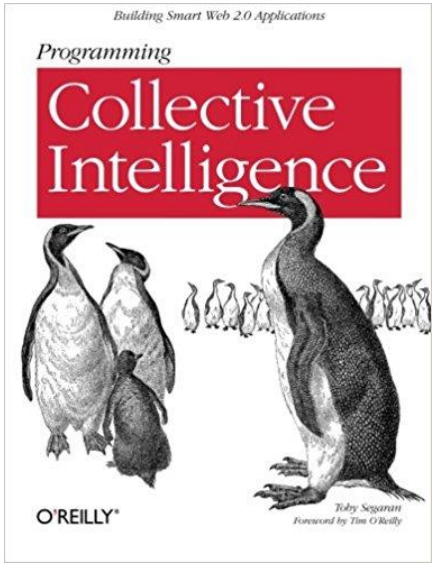
$29.88

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Books



Machine Learning: An Algorithmic Perspective
by Stephen Marsland



Programming Collective Intelligence
by Toby Segaran

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Types of assignments

- Toy problem solving – at home every week. Required for the next class –absolutely no late submissions 10%
- Weekly labs – real coding, real problems, smart applications that learn: 30%
- 2 midterm tests 20%
- Final project: open-ended: 40%

- What is machine learning
- Why ML
- Types of ML tasks
- Course requirements

Labs: learning by doing

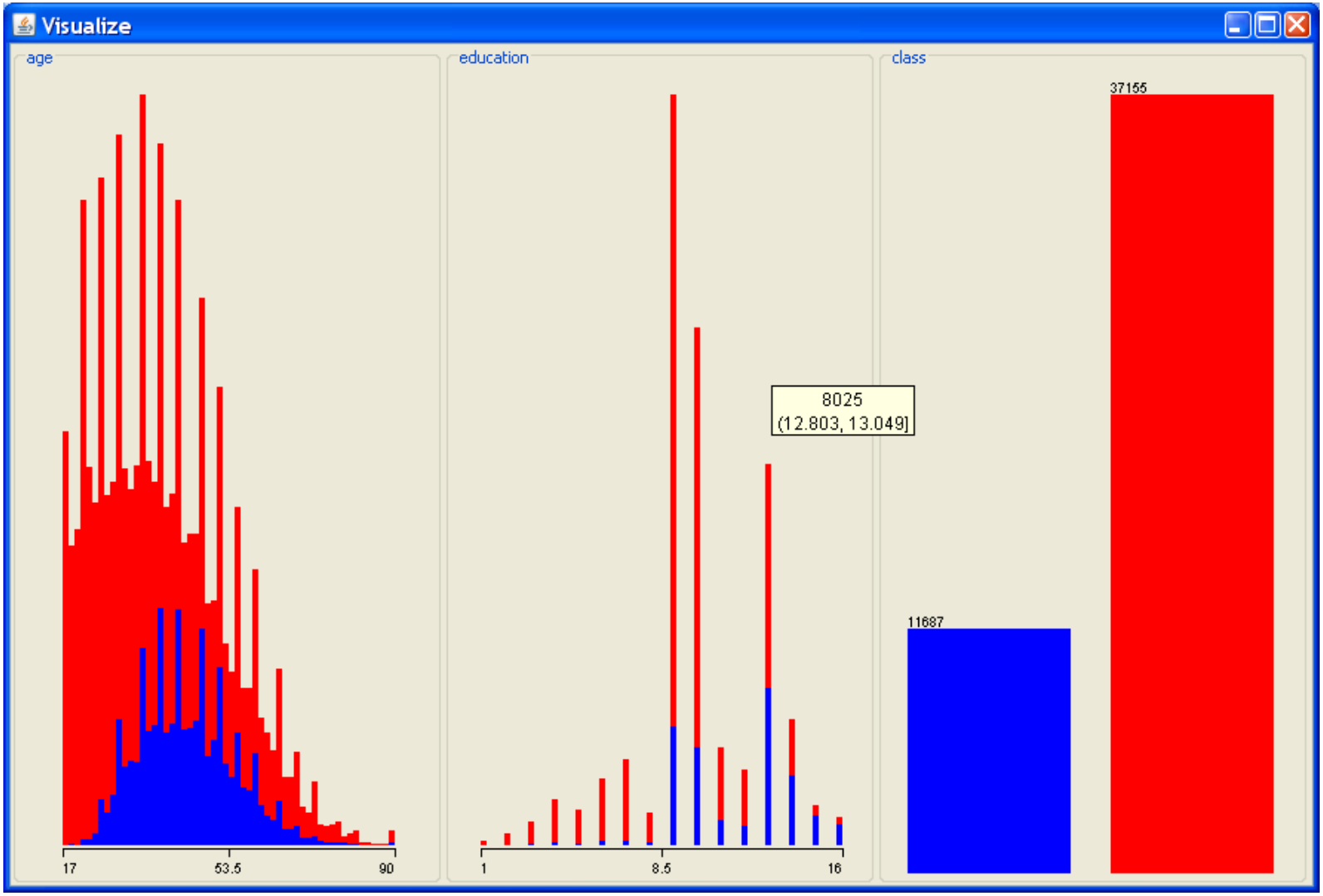
- Learning by example: on toy datasets which exhibit features of real-life datasets
- Python implementation of ML algorithms
- Real-life datasets analysis

Lab example: what determines high salary

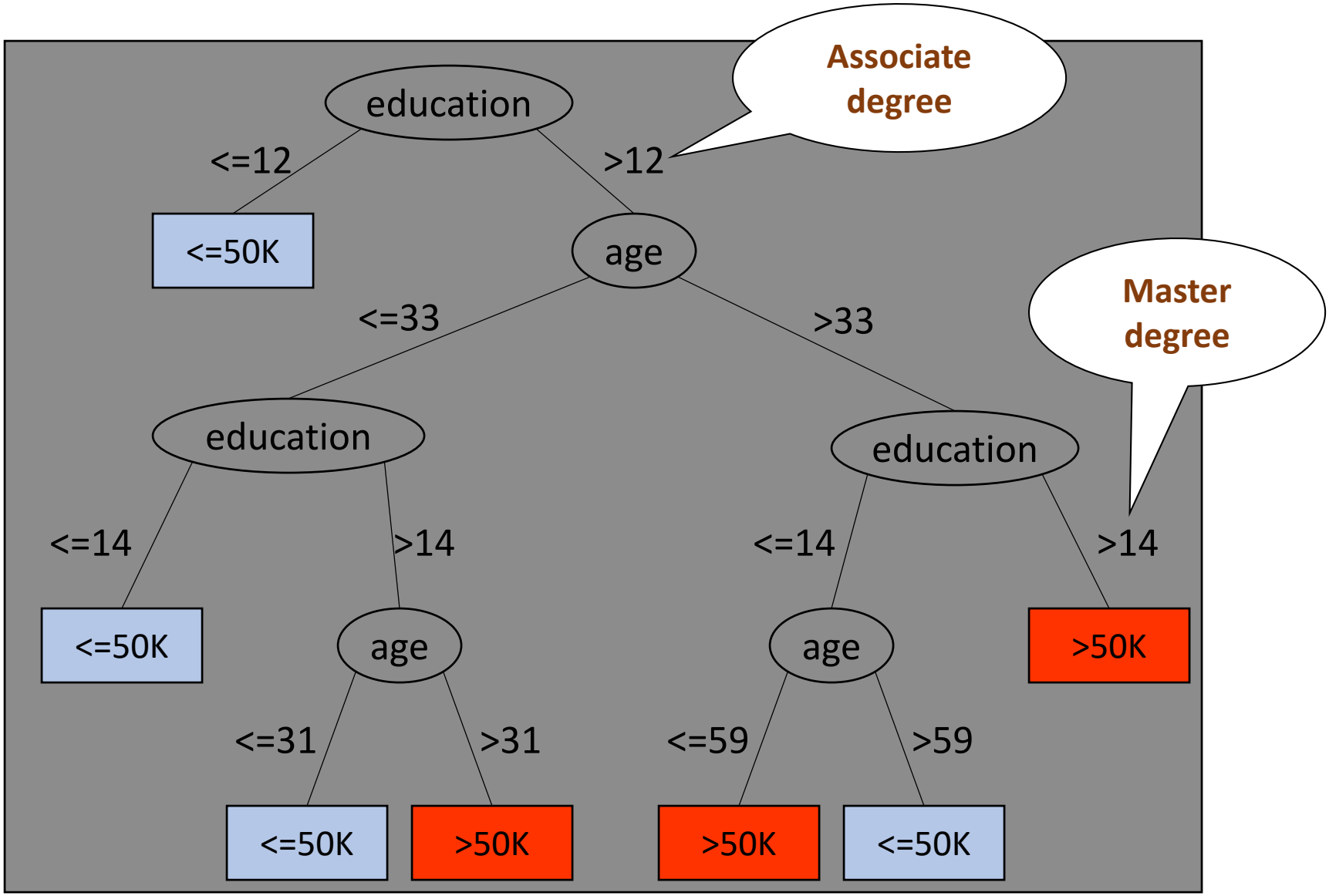
Adult income dataset (US census 1994)

Age	Education	Mar. status	Occupation	Race	Sex	Born in	Yearly income
39	Bachelors	Never-married	Adm-clerical	White	M	US	**<=50 K**
50	Bachelors	Married-civ-spouse	Exec-managerial	White	M	US	**<=50 K**
54	7th-8th	Married-civ-spouse	Machine-op-inspct	White	M	US	**>50K**
37	Bachelors	Never-married	Exec-managerial	Black	M	US	**>50K**
28	Bachelors	Married-civ-spouse	Prof-specialty	Black	F	Cuba	**<=50 K**
37	Masters	Married-civ-spouse	Exec-managerial	White	F	US	**<=50 K**

Visualization of attributes: *age* and *education*



The results of learning: decision tree on age and education attributes



Course syllabus

I. Supervised learning

1. **Decision trees**.
   Information theory, entropy, information gain, variance
2. **Probabilistic classifiers: Naive Bayes and Bayesian Belief Networks.**
   Belief and probability, conditional prob
3. **Evaluation of learning.**
   Confidence intervals, credibility, T-test
   Signal theory: ROC curves
4. **Instance-based learning: k-Nearest Neighbors.**
   Similarity and distance, scaling

- **II. Unsupervised learning**
  5. **Associations and correlations.**
     Correlation measures
  6. **Clustering, fuzzy clustering.**
     Expectation maximization
- **III. Advanced Supervised learning**
  7. **Regression, logistic regression.**
     Linear algebra, least squares
  8. **Separators: Support Vector Machines**
     Hyperplane geometry, perceptron
  9. **Artificial Neural Networks**
- **IV. Advanced topics**
  10. **Dimensionality reduction.**
      Principal Component Analysis. Singular Value Decomposition.
  11. **Network learning: PageRank.**
  12. **Genetic algorithm.**