203.4770: Introduction to Machine Learning

Dr. Rita Osadchy

Outline

1. About the Course
2. What is Machine Learning?
3. Types of problems and Situations
4. ML Example

About the course

- Course Homepage:
  http://www.cs.haifa.ac.il/~rita/ml_course/course.html
- Office hours: request meeting by email
- Contact:
  - You contact me by email: rita@cs.haifa.ac.il
  - I contact you by email: All announcement, home assignments, and guidelines will be distributed by email.

You must send me an email by November 7 from your active address with the subject “ML course contact”.

  - Those who do not send their contact address on time will not be added to the contact list!!!

Prerequisites

- The course assumes some basic knowledge of the probability theory, linear algebra, and basic programming skills.
- You should be familiar with:
  - Joint and marginal probability distributions
  - Normal (Gaussian) distribution
  - Expectation and variance
  - Statistical correlation and statistical independence

Probability/ Statistics

  - Matrices, vectors, and their multiplication
  - Matrix inverse
  - Eigen value decomposition

Linear Algebra

Links to tutorial in the course homepage.

Course Material:

- **Textbooks:**
  - Duda, R. O. Hart, P. E. D., and Stork, G.*Pattern Classification .* New York, NY: Wiley, 2000.
  - T. Hastie, R. Tibshirani, and J. Friedman: "Elements of Statistical Learning", Springer-Verlag, 2001.
  - Pattern Recognition and Machine Learning, by Christopher Bishop. Springer, August 2006.
- lecture notes and reading material in: http://www.cs.haifa.ac.il/~rita/ML_course/course.htm

Final Grade

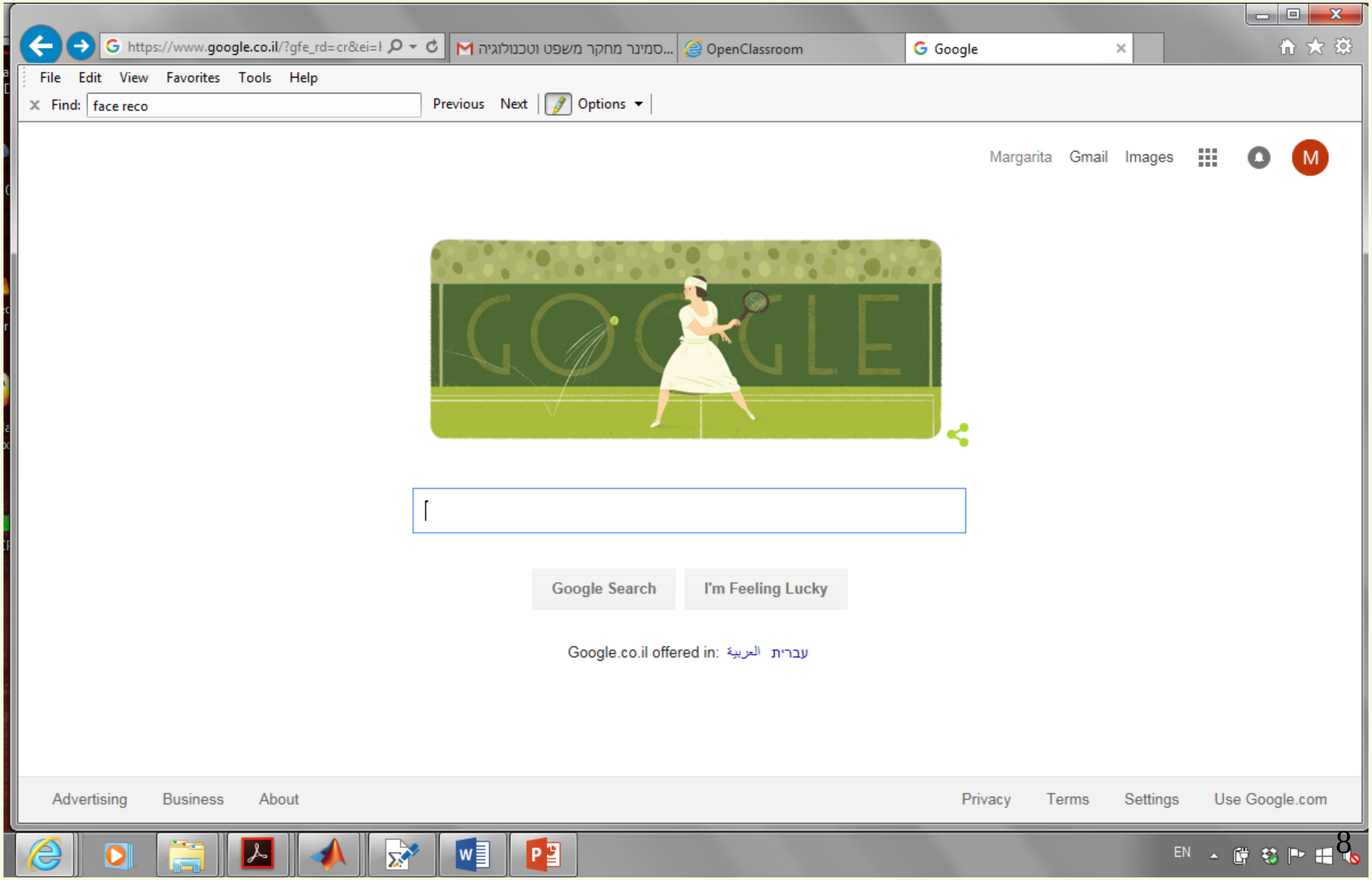
- Home assignments (depending on the number of students)
  - 0%-20%
- Final Exam
  - 80%-100%

Home Assignments

- Mostly practical (implement learning algorithms)
- Programming in Matlab
  - Very easy to write code that operates matrices and use plots.
- Submission in pairs (not allowed to change groups after the first assignment)
- Discussions between groups are allowed, same solutions are not allowed!

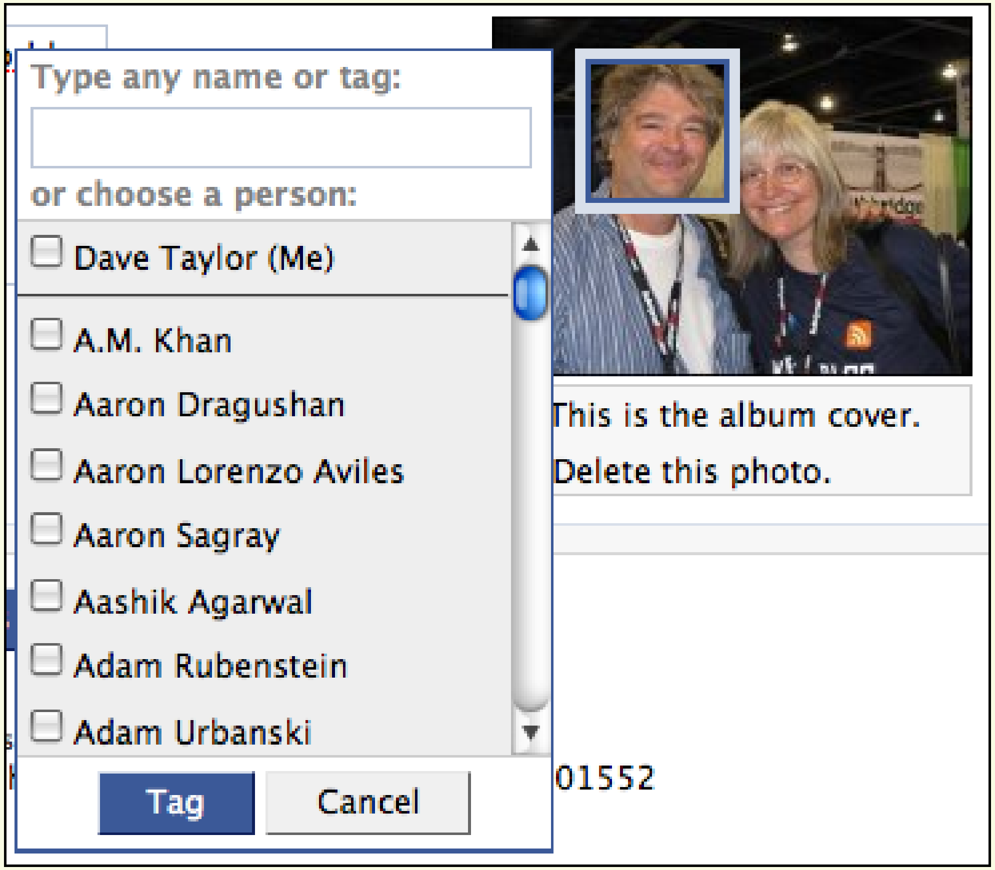
You used ML today…

Google search



You used ML today…

Tagging photos on Facebook or mobile device



You used ML today…

Spam Filters

Why do we need Machine Learning?

Database mining – large datasets from growth of automation/web:

- Web click data
- Medical records
- Biology
- Engineering
- Finance

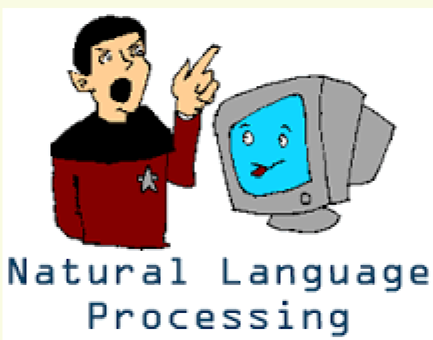
Why do we need Machine Learning?

Applications that cannot be explicitly programmed.

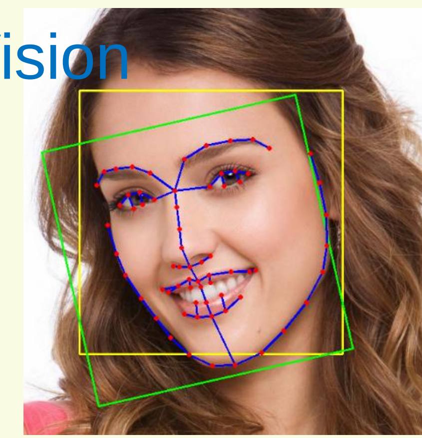
OCR

Computer Program

Thank you



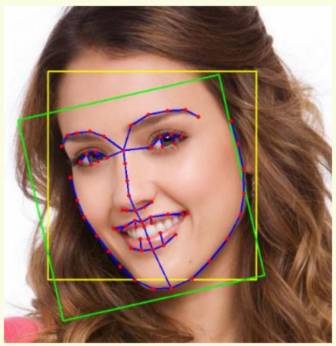
Computer Vision



Robot Navigation

Why do we need Machine Learning?

Applications that cannot be explicitly programmed.



Self-driving car

Why do we need Machine Learning?

Self-customizing programs

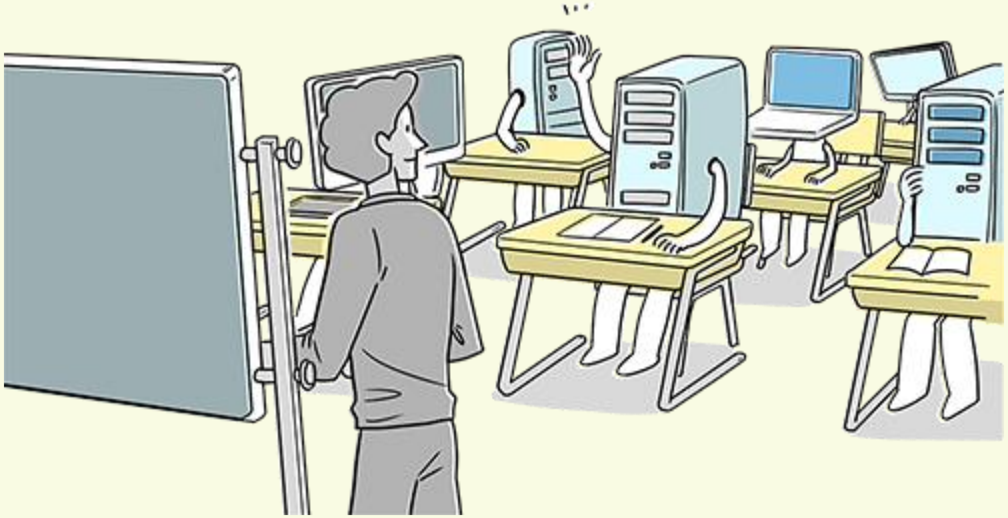
Product recommendations

What is Learning?

- Learning is an essential human property
- Learning: Acquisition of knowledge, understanding, and ability with experience.
- Learning IS NOT learning by heart

What is Machine Learning?

- Any computer can learn by heart, the difficulty is to make a prediction – generalize a behavior to a novel situation.



Machine Learning - Definition

Study of algorithms that improve their performance P at some task T with experience E

OCR

T: recognition of a handwritten letter “a” from its image.

E: images of a handwritten “a”.

P: recognition rate.

Types of Machine Learning Problems

- Supervised learning:
  The correct answers are given

- Unsupervised learning:
  Find structure in the world

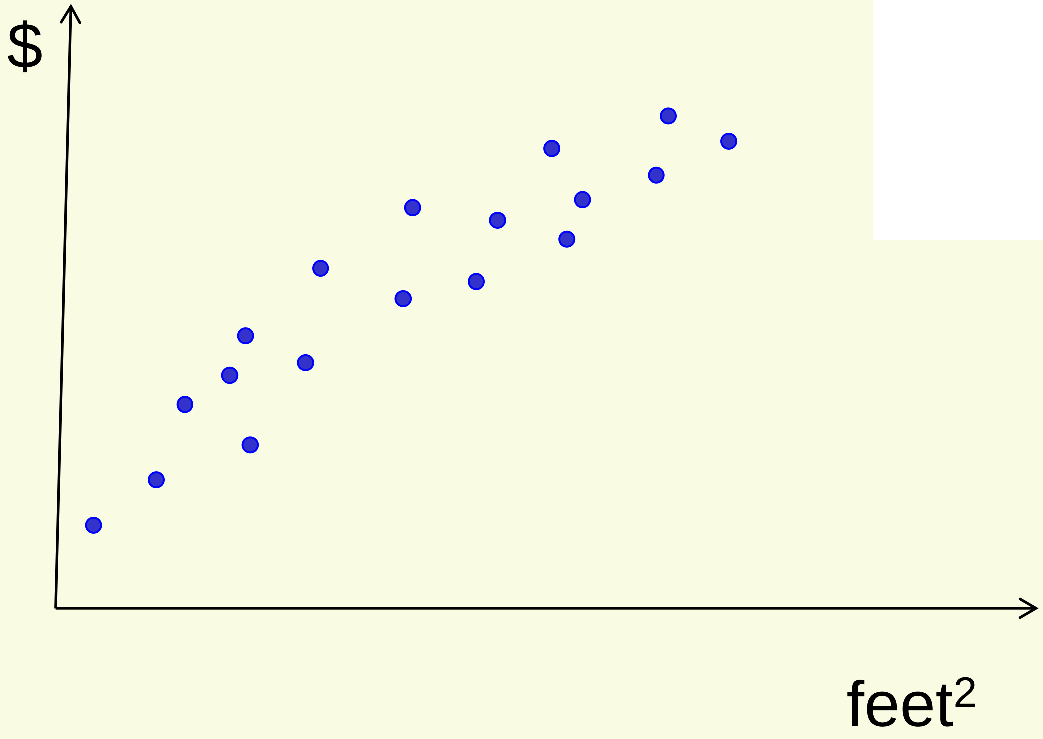
- Other: reinforcement learning, recommender systems…

Supervised Learning

Given a set of training inputs and corresponding outputs (correct answers), produce the “correct” outputs for the new inputs.

Regression

- Housing Prices



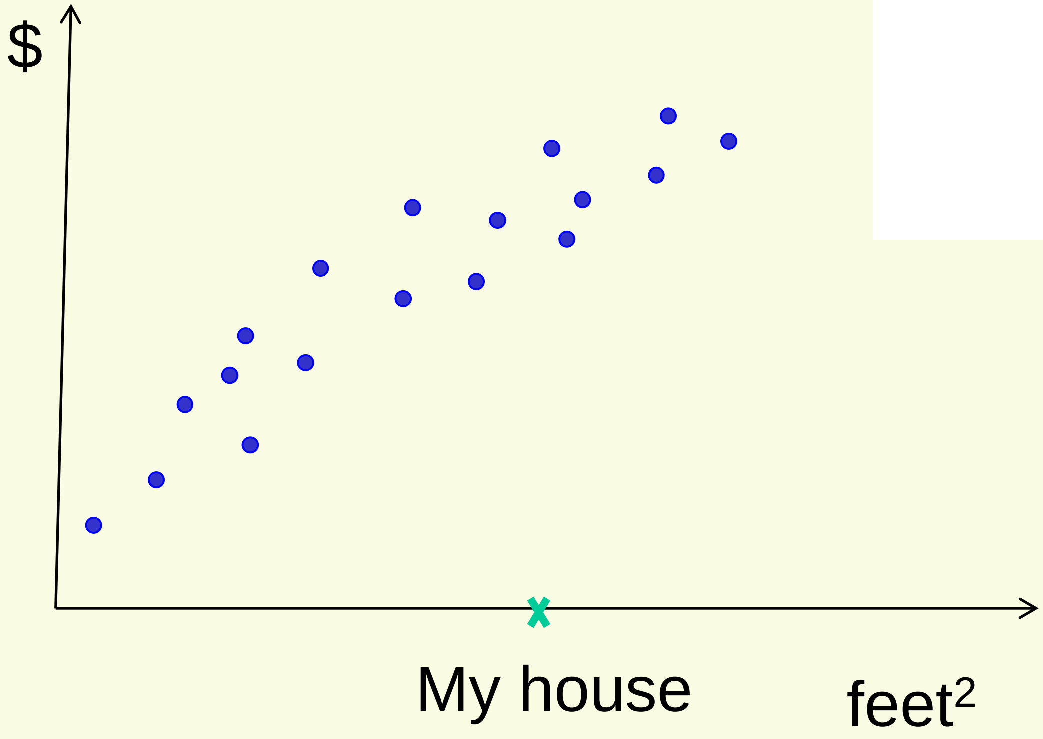
Living area (feet^2)	Price (1000$s)
2104	400
1600	330
2400	369
1416	232
3000	540
⋮	⋮

Example by Andrew Ng, Stanford

Regression

- Housing Prices

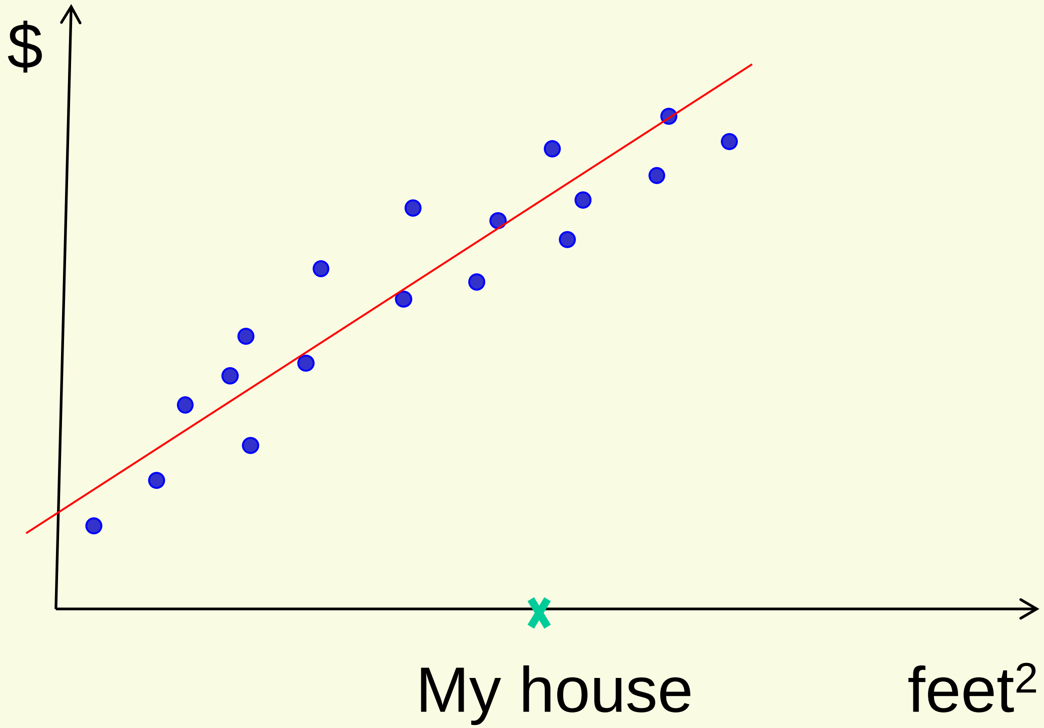
Living area ($feet^2$)	Price (1000\$s)
2104	400
1600	330
2400	369
1416	232
3000	540
⋮	⋮



Example by Andrew Ng, Stanford

Regression

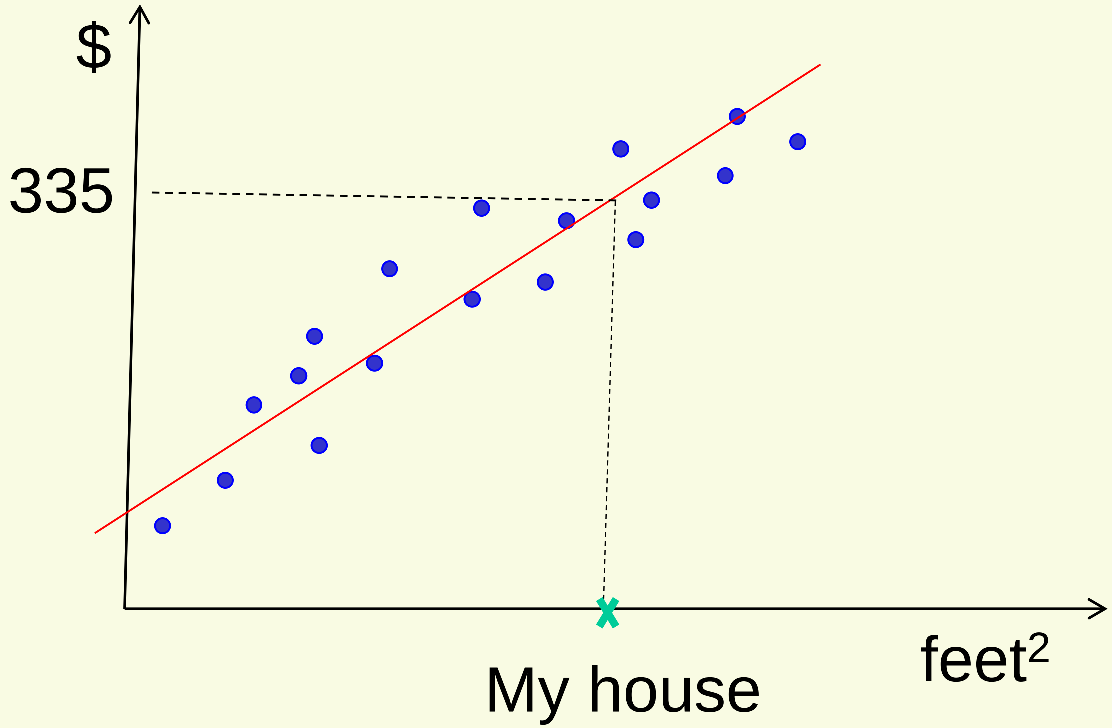
- Housing Prices



Example by Andrew Ng, Stanford

Regression

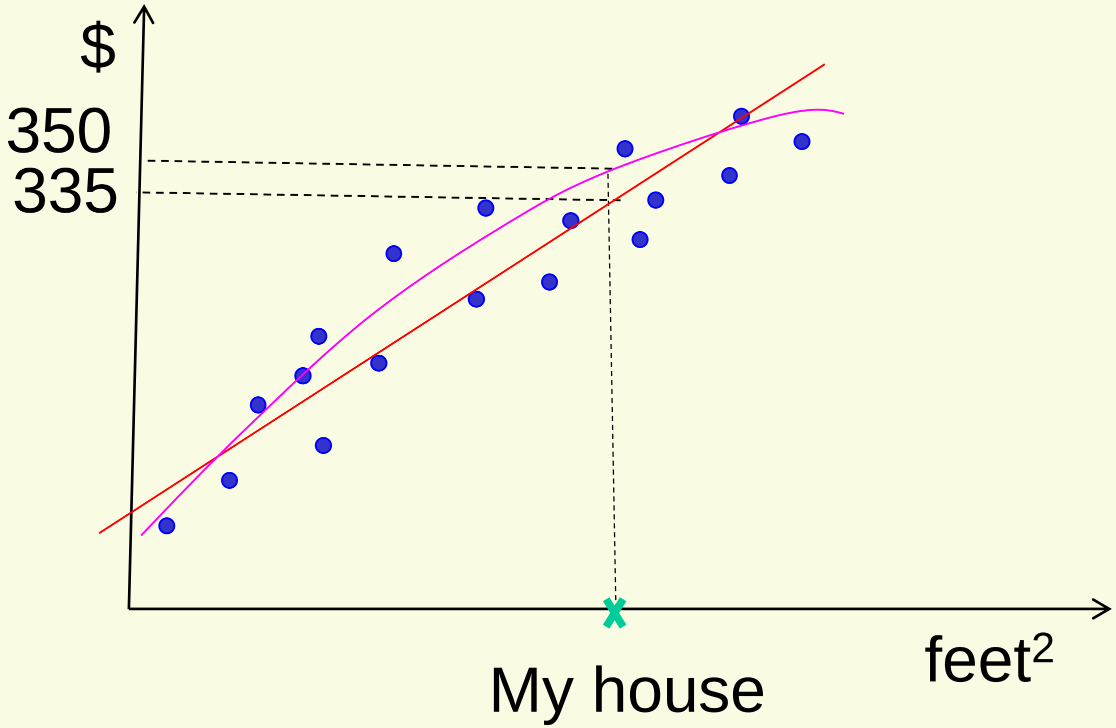
- Housing Prices



Example by Andrew Ng, Stanford

Regression

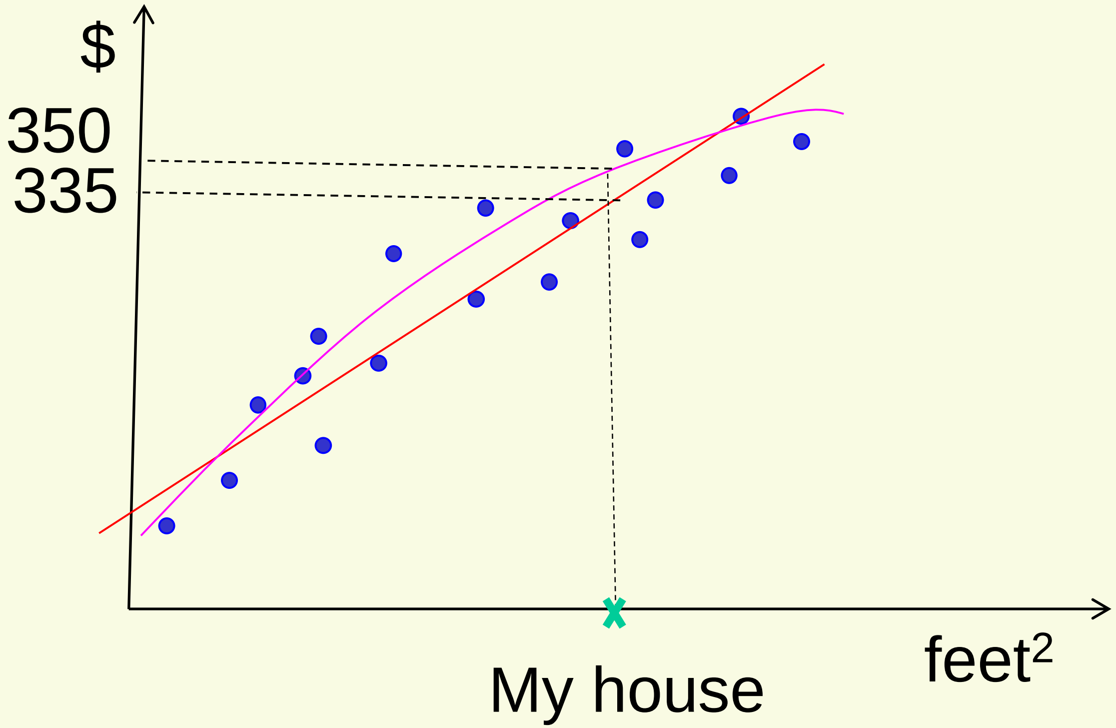
- Housing Prices



Regression

- Housing Prices

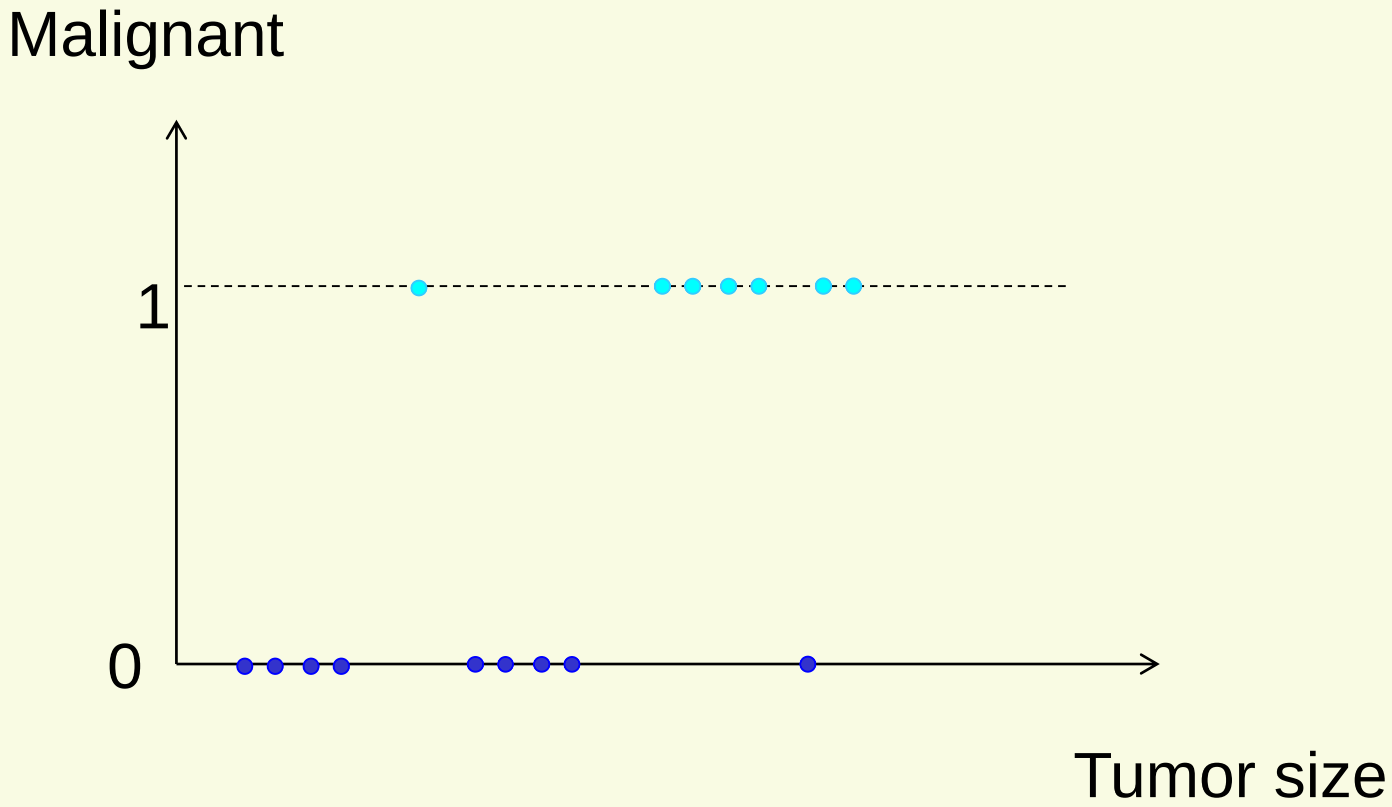
Regression – continues output



Example by Andrew Ng, Stanford

Classification

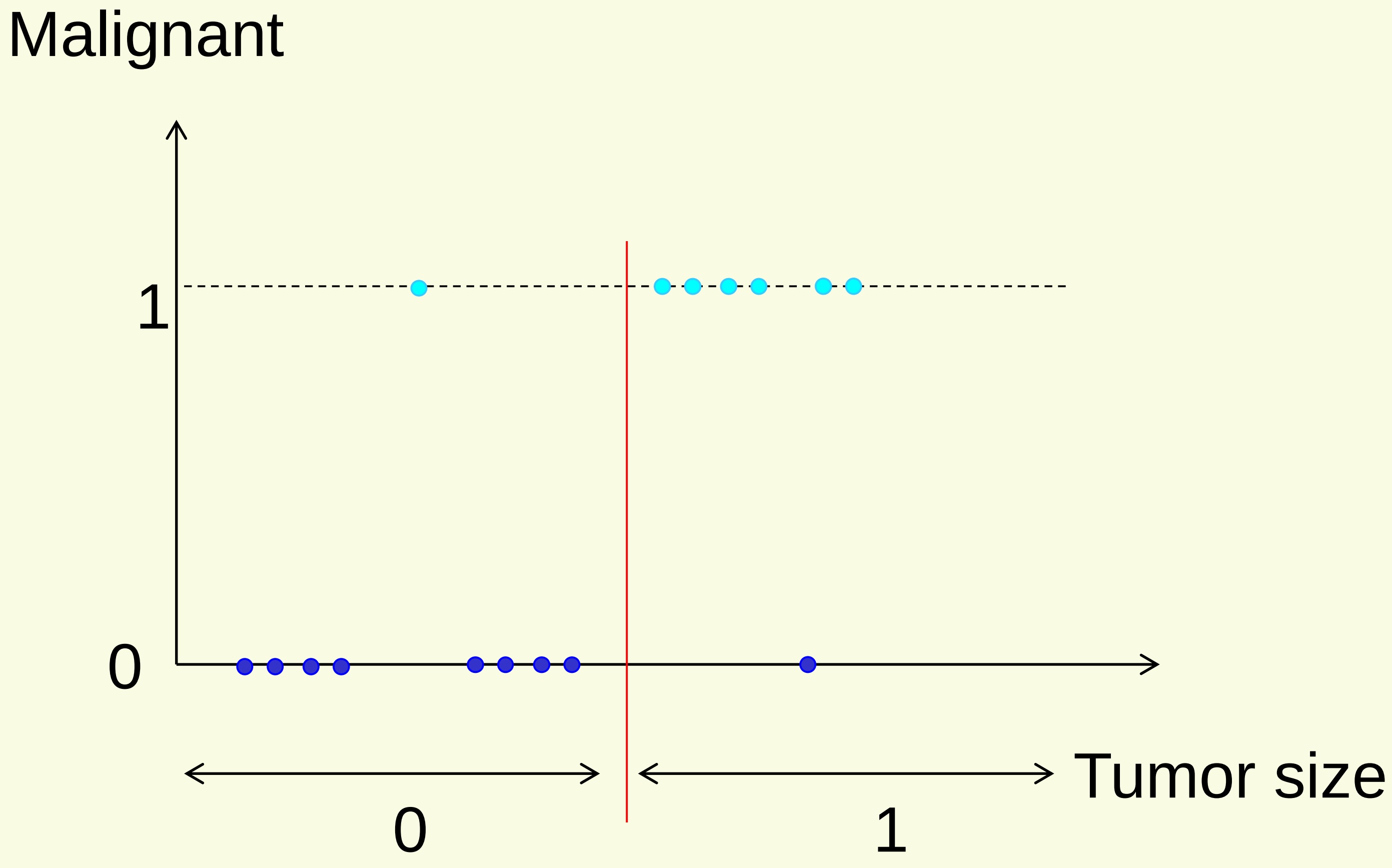
- Tumor Classification



Example by Andrew Ng, Stanford

Classification

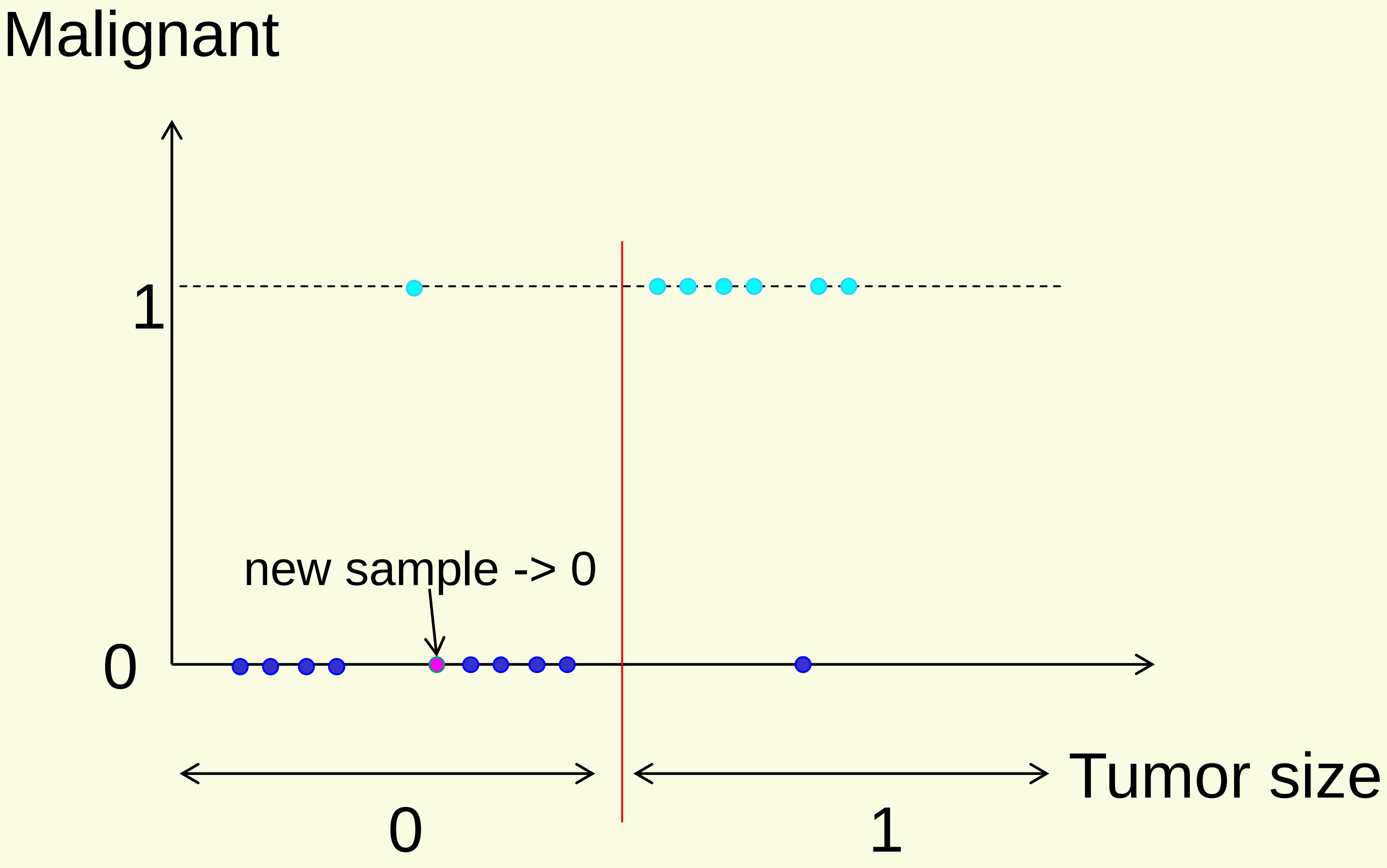
- Tumor Classification



Example by Andrew Ng, Stanford

Classification

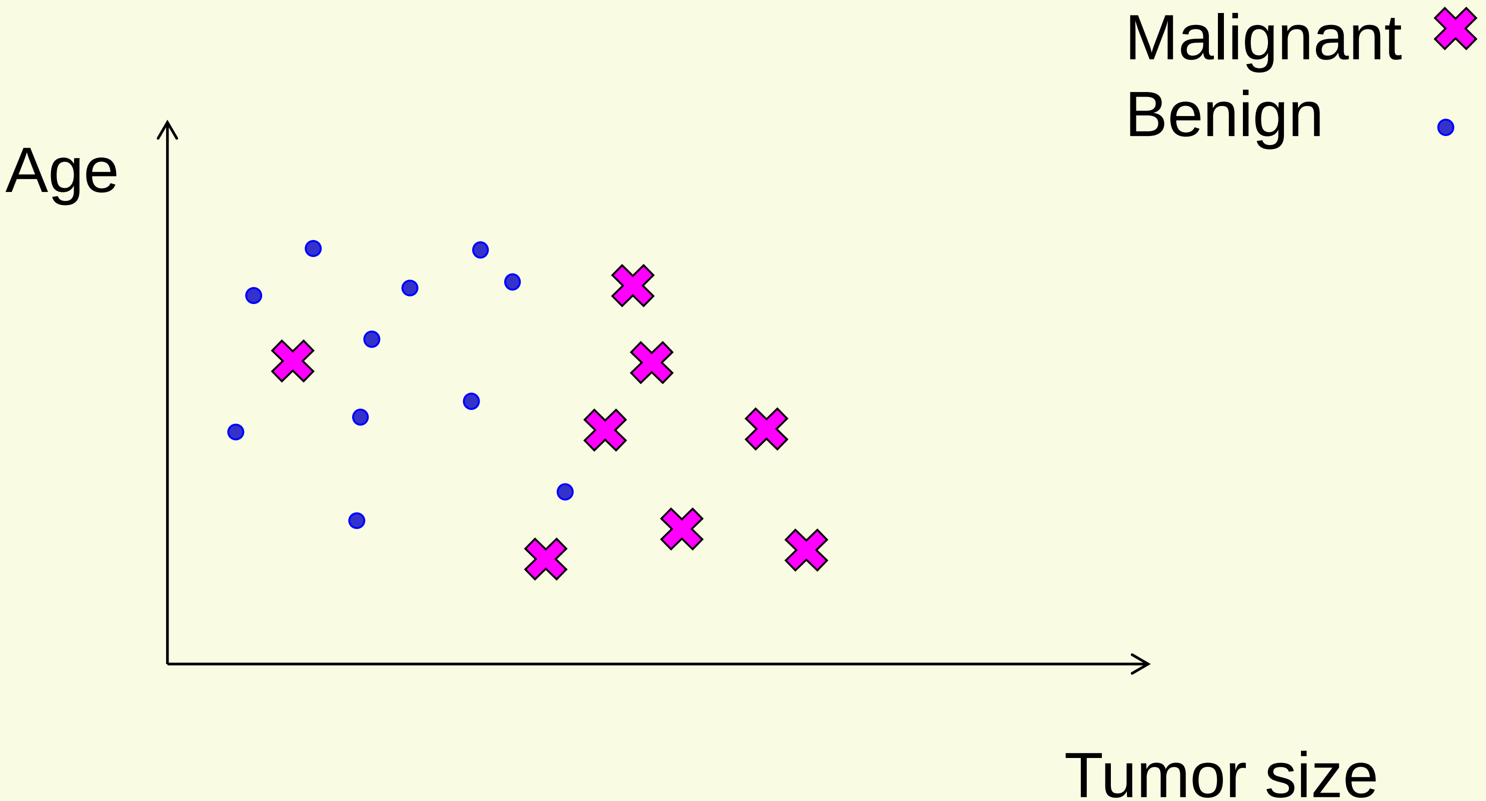
- Tumor Classification



Example by Andrew Ng, Stanford

Classification

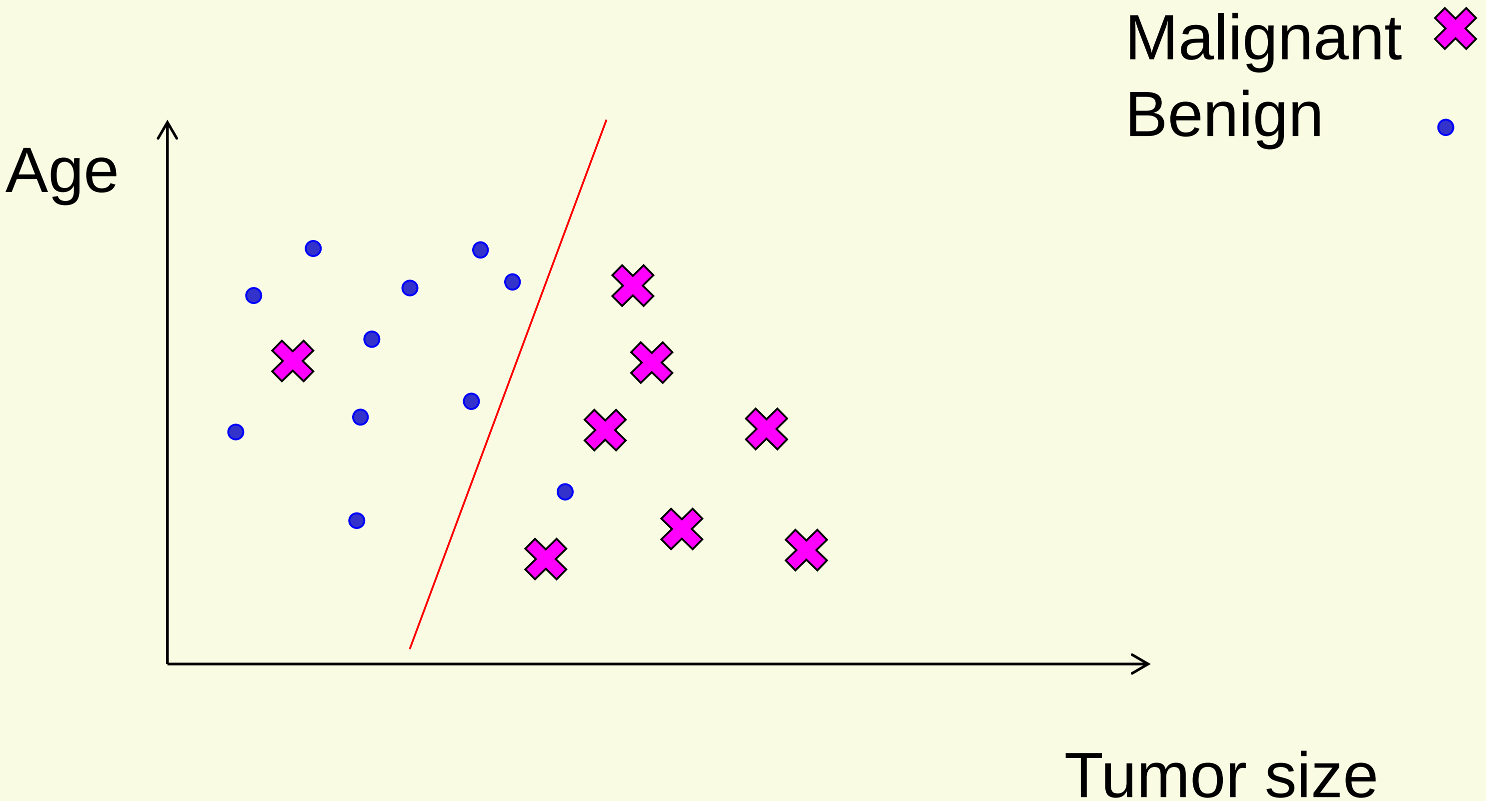
- Tumor Classification



Example by Andrew Ng, Stanford

Classification

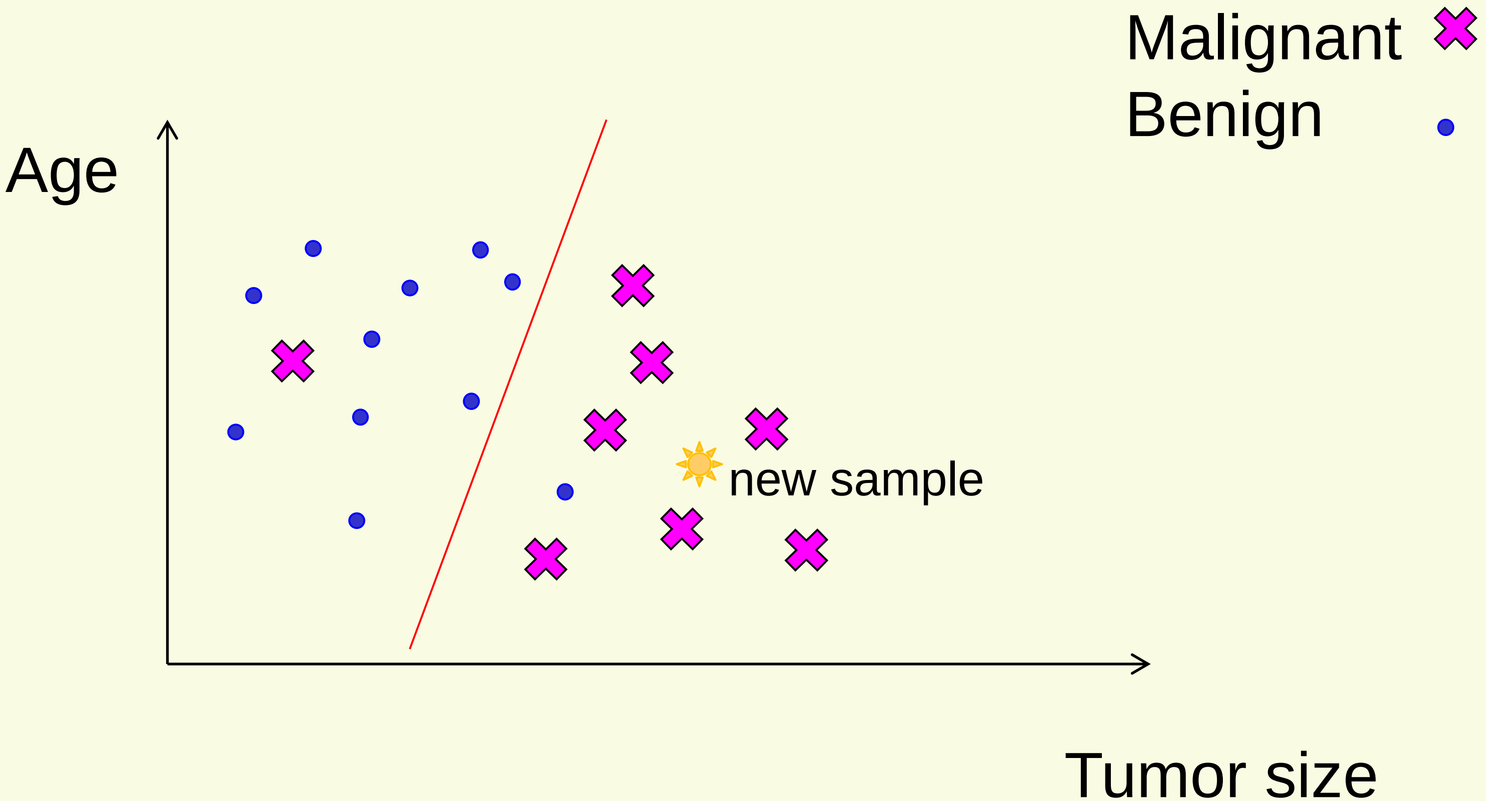
- Tumor Classification



Example by Andrew Ng, Stanford

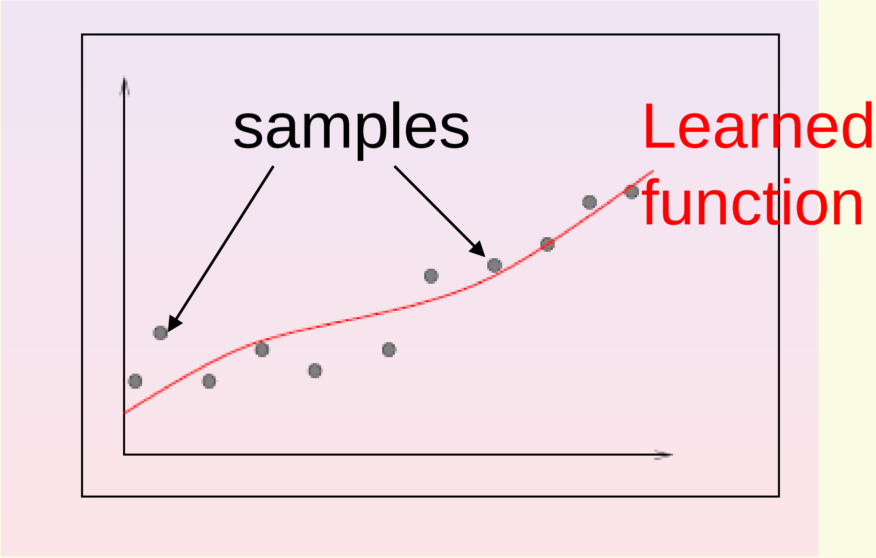
Classification

- Tumor Classification

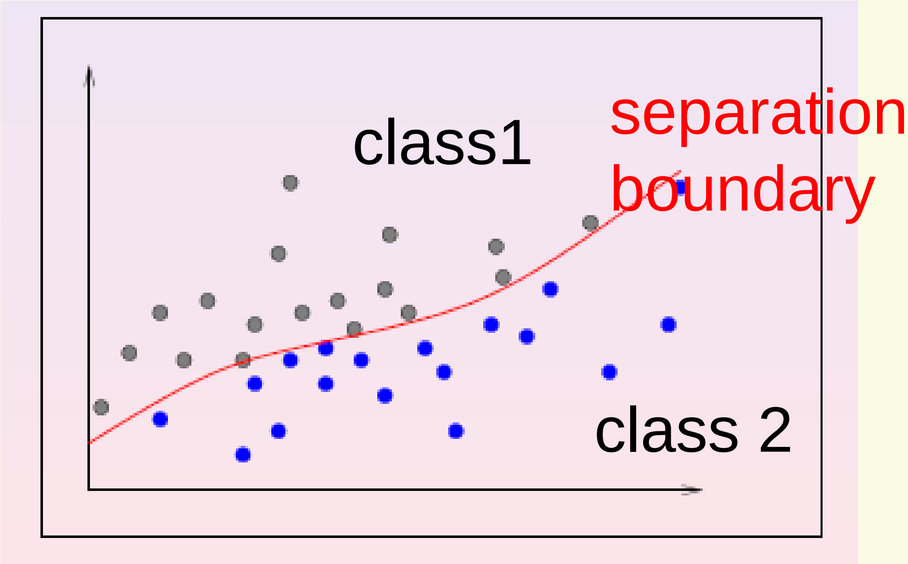


Example by Andrew Ng, Stanford

Two kinds of Supervised Learning

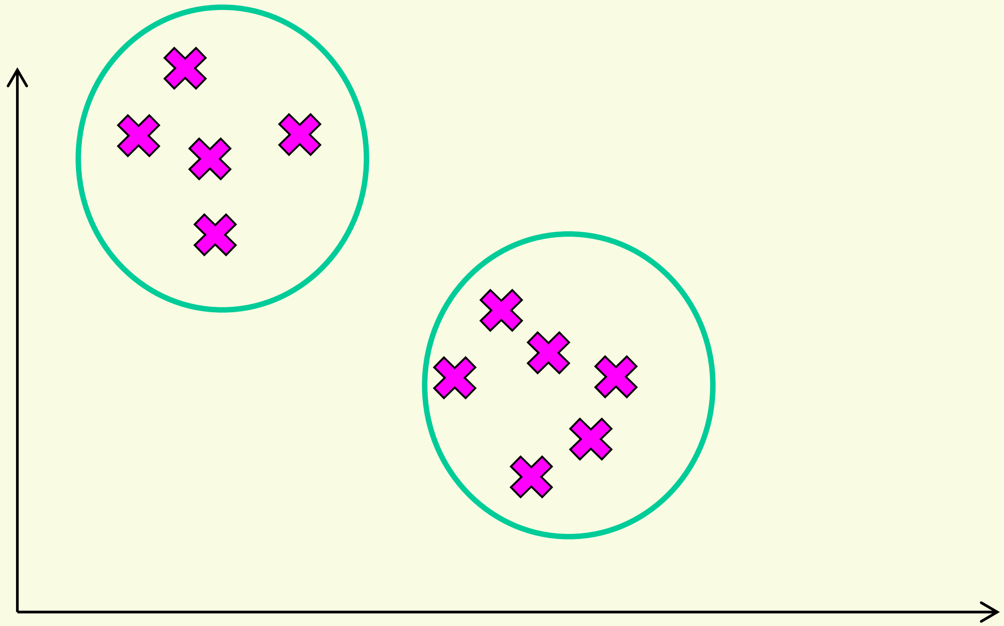


- Regression: Learn a continuous input-output mapping from a limited number of examples.



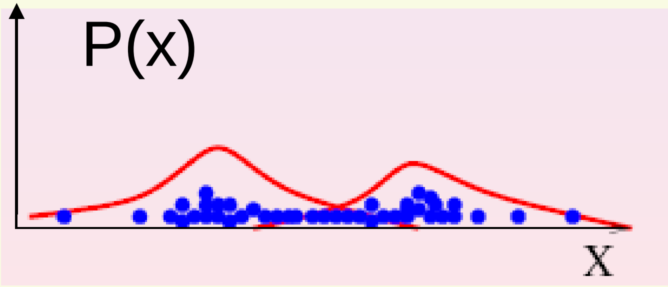
- Classification: outputs are discrete variables (category labels). Learn a decision boundary that separates one class from the other.

Unsupervised learning

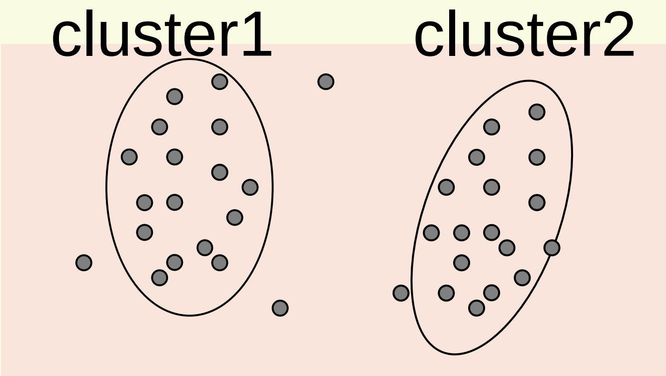


Given only inputs as training, find structure in the world: discover clusters, manifolds, characterize the areas of the space to which the observed inputs belong.

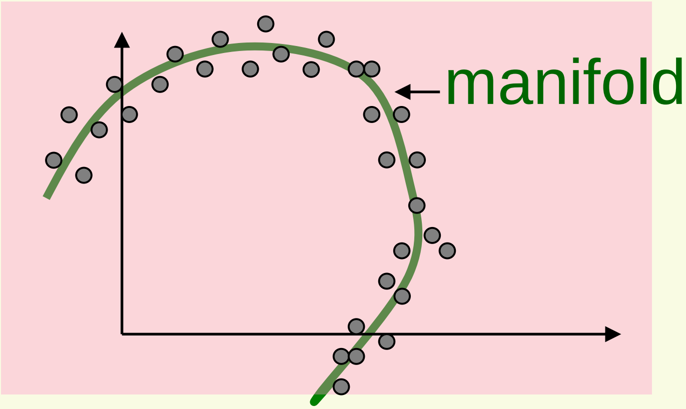
Unsupervised Learning



- Density Estimation. Find a function f such f(X) approximates the probability density of X, p(X), as well as possible.



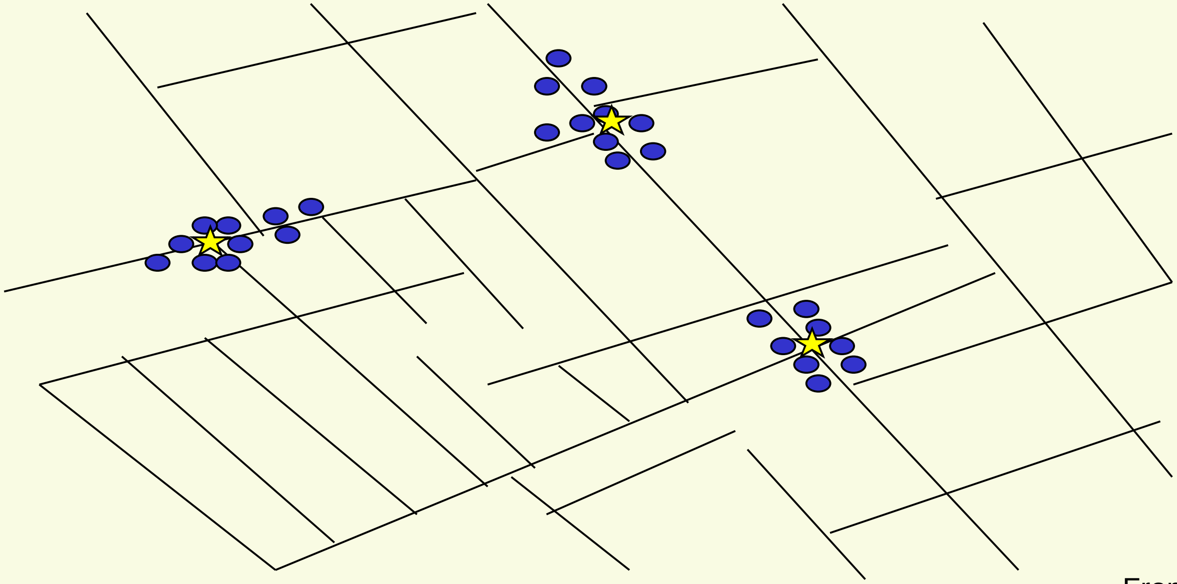
- Clustering: discover “clumps” of points



- Embedding: discover low-dimensional manifold or surface near which the data lives.

First (?) Application of Clustering

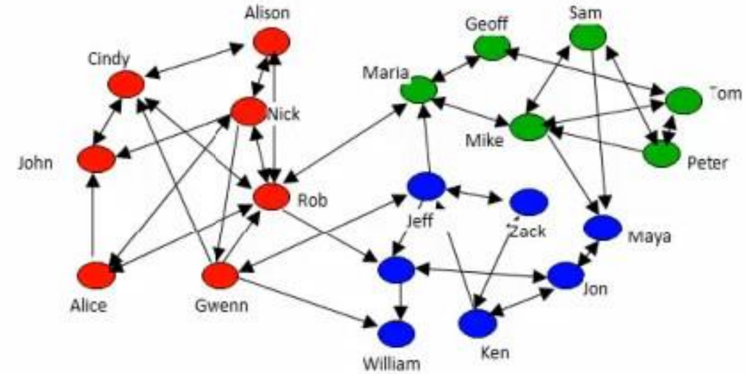
- John Snow, a London physician plotted the location of cholera deaths on a map during an outbreak in the 1850s.
- The locations indicated that cases were clustered around certain intersections where there were polluted wells -- thus exposing both the problem and the solution.



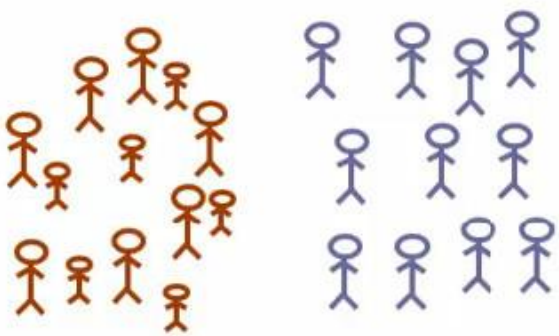
From: Nina Mishra HP Labs

Clustering Applications

Organize computing clusters



Social network analysis



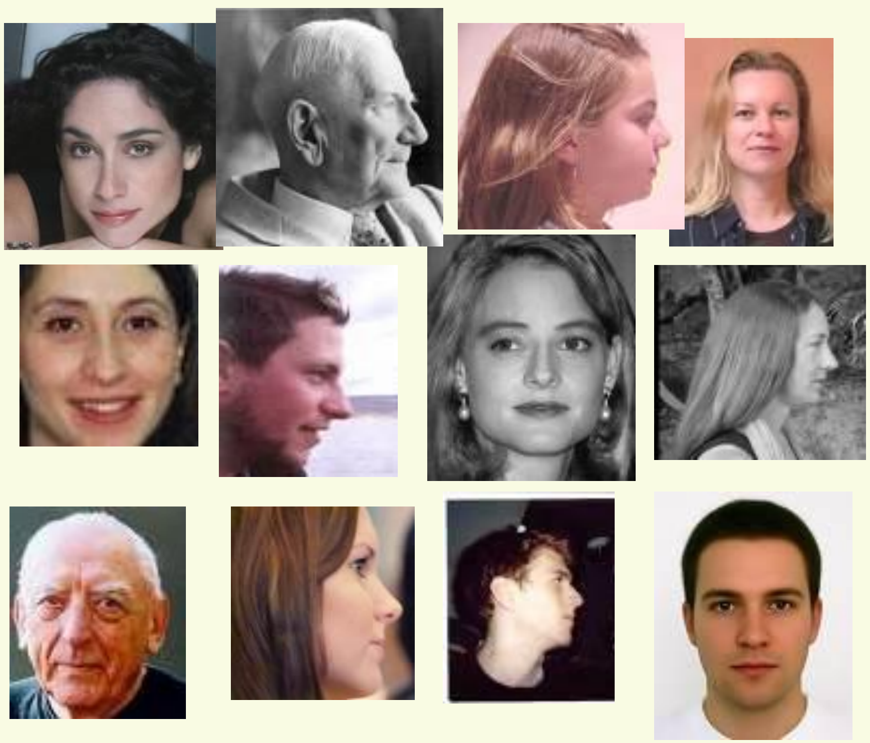
Market segmentation.

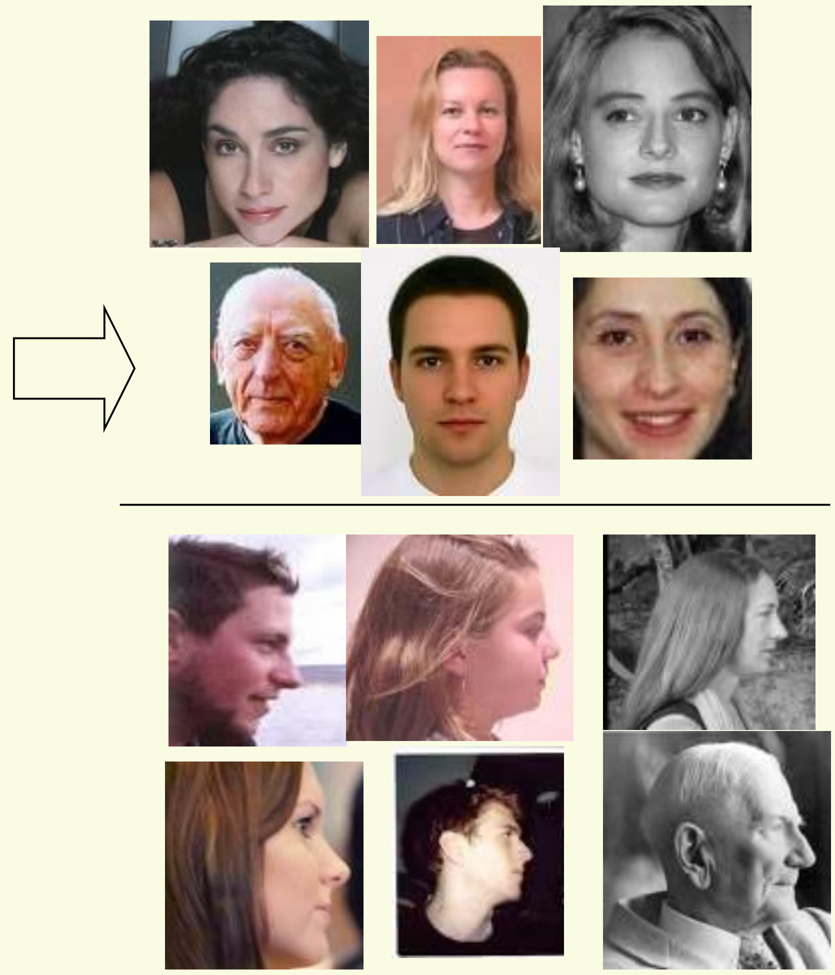
Astronomical data analysis

Andrew Ng

Clustering Example

- Cluster images of faces into two groups

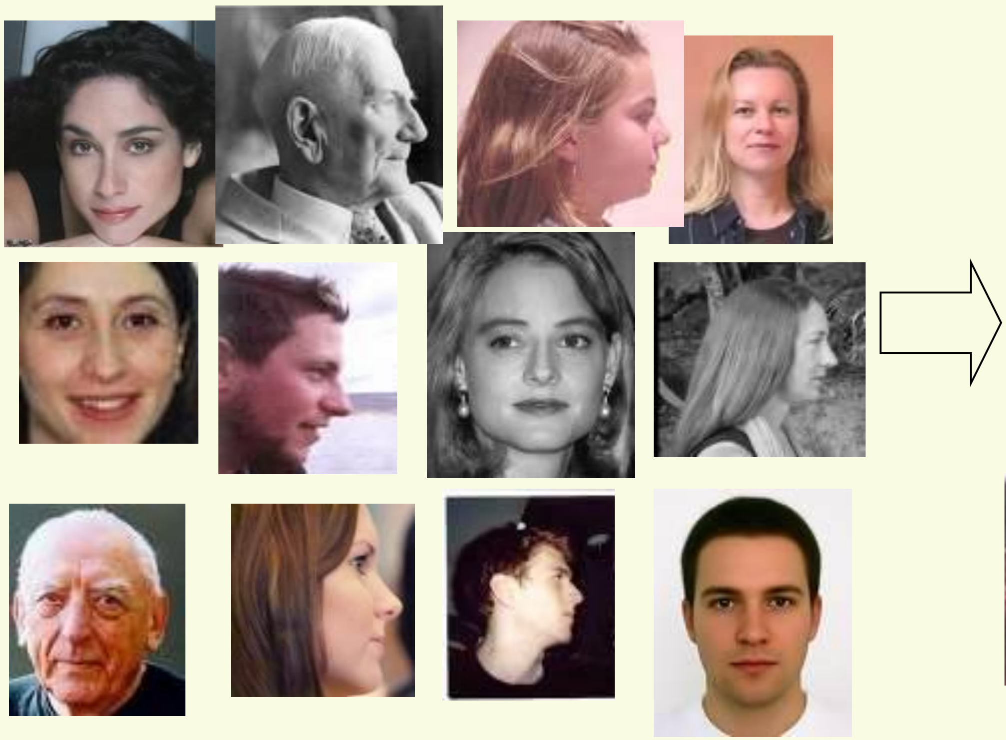


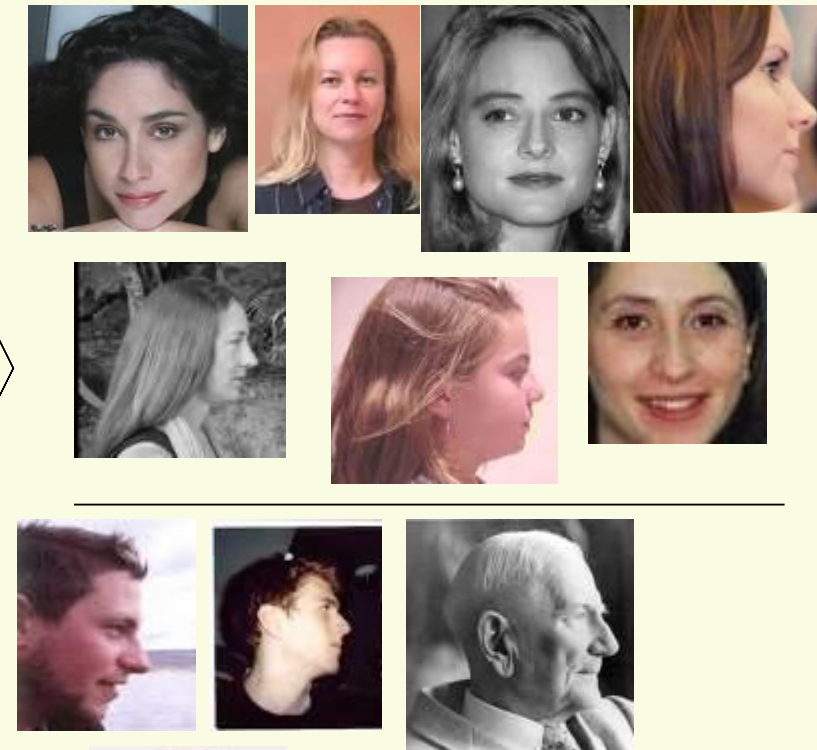


by view point

Clustering Example

- Cluster images of faces into two groups





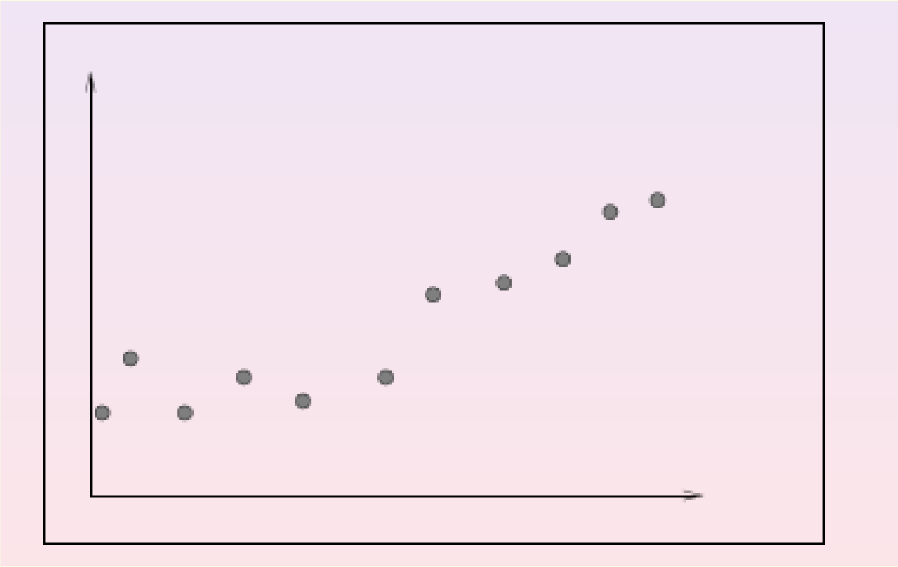
by gender

Types of Learning Problems

- Reinforcement learning, where we only get feedback in the form of how well we are doing (For example the outcome of the game).
  - Don’t have time to discuss in this course ☹

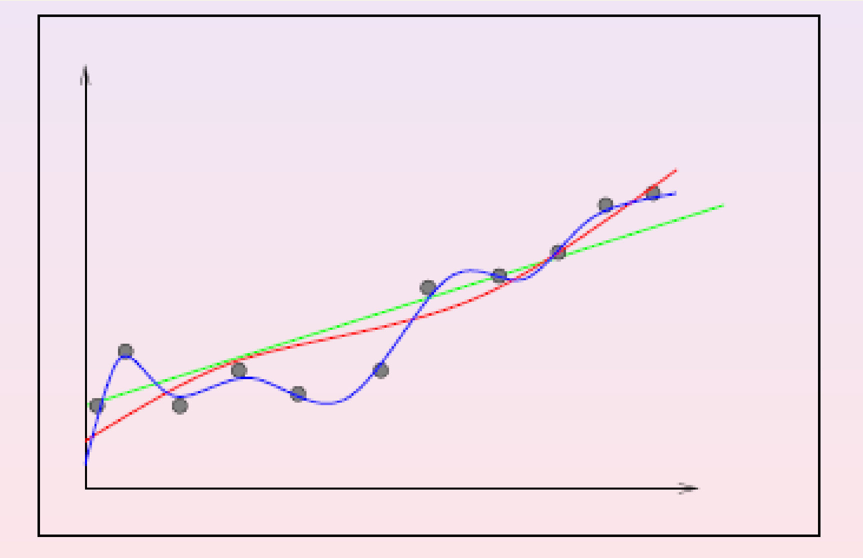
Why Learning is Difficult?

- Given a finite amount of training data, you have to derive a relation for an infinite domain.
- In fact, there is an infinite number of such relations



Why Learning is Difficult?

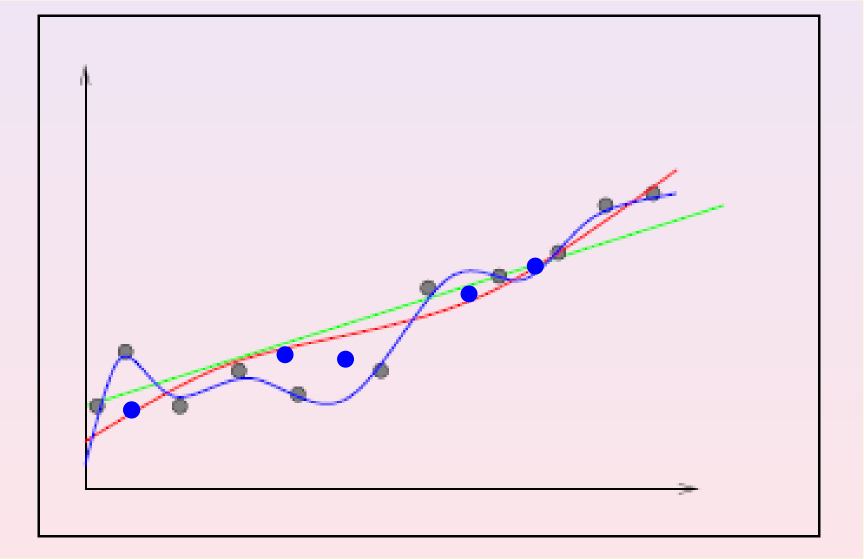
- Given a finite amount of training data, you have to derive a relation for an infinite domain
- In fact, there is an infinite number of such relations



- Which relation is more appropriate?

Why Learning is Difficult?

- Given a finite amount of training data, you have to derive a relation for an infinite domain
- In fact, there is an infinite number of such relations



- ... the hidden test points...

Occam's Razor's Principle

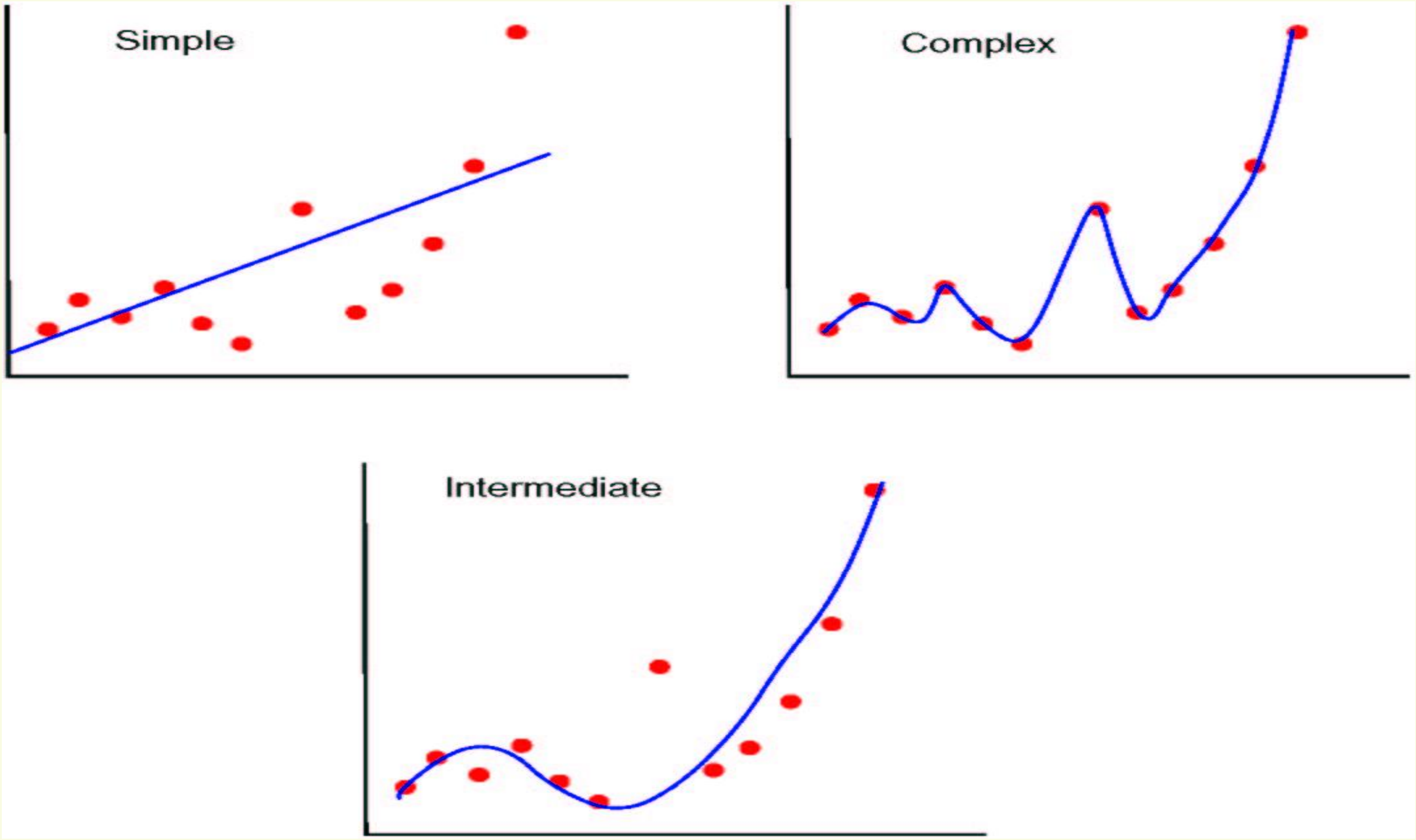
- Occam's Razor's Principle(14th century):

  One should not increase, beyond what is necessary, the number of entities required to explain anything

- When many solutions are available for a given problem, we should select the simplest one.
- But what do we mean by simple?
- We will use prior knowledge of the problem to define what is a simple solution.

Example of a prior: smoothness

Generalization in Regression



Example

- A classification problem: predict the grades for students taking this course.

Slide credit: Tommi Jaakkola, MIT

Example

- A classification problem: predict the grades for students taking this course.
- Key steps:
  - Data: what “past experience” can we rely on?

Slide credit: Tommi Jaakkola, MIT

Example

- A classification problem: predict the grades for students taking this course.
- Key steps:
  - Data: what “past experience” can we rely on?
  - Assumptions: what can we assume about the students or the course?

Slide credit: Tommi Jaakkola, MIT

Example

- A classification problem: predict the grades for students taking this course.
- Key steps:
  - Data: what “past experience” can we rely on?
  - Assumptions: what can we assume about the students or the course?
  - Representation: how do we “summarize” a student?

Slide credit: Tommi Jaakkola, MIT

Example

- A classification problem: predict the grades for students taking this course.
- Key steps:
  - Data: what “past experience” can we rely on?
  - Assumptions: what can we assume about the students or the course?
  - Representation: how do we “summarize” a student?
  - Estimation: how do we construct a map from students to grades?

Slide credit: Tommi Jaakkola, MIT

Example

- A classification problem: predict the grades for students taking this course.
- Key steps:
  - Data: what “past experience” can we rely on?
  - Assumptions: what can we assume about the students or the course?
  - Representation: how do we “summarize” a student?
  - Estimation: how do we construct a map from students to grades?
  - Evaluation: how well are we predicting?

Slide credit: Tommi Jaakkola, MIT

Example

- A classification problem: predict the grades for students taking this course.
- Key steps:
  - Data: what “past experience” can we rely on?
  - Assumptions: what can we assume about the students or the course?
  - Representation: how do we “summarize” a student?
  - Estimation: how do we construct a map from students to grades?
  - Evaluation: how well are we predicting?
  - Model selection: perhaps we can do even better?

Slide credit: Tommi Jaakkola, MIT

Data

- The data we have available (in principle):
  - names and grades of students in past years ML courses
  - academic record of past and current students
- “training" data:

Student	ML	course1	course2	…
Peter	A	B	A	…
David	B	A	A	…

- “test" data:

Student	ML	course1	course2	…
Jack	?	C	A	…
Kate	?	A	A	…

- Anything else we could use?

Slide credit: Tommi Jaakkola, MIT

Assumptions

- There are many assumptions we can make to facilitate predictions
  1. the course has remained roughly the same over the years
  2. each student performs independently from others

Slide credit: Tommi Jaakkola, MIT

Presentation

- Academic records are rather diverse so we might limit the summaries to a select few courses
- For example, we can summarize the *i*th student (say Pete) with a vector

$$\mathrm{x}_i = [100 \;\; 60 \;\; 80]$$

- The available data in this representation

Training		Test	
Student	ML grade	Student	ML grade
x_1	100	x'_1	?
x_2	80	x'_2	?
...	...	...	...

Slide credit: Tommi Jaakkola, MIT

Estimation

- Given the training data
  we need to find a mapping from “input vectors” x to “labels“ y encoding the grades for the ML course.

Student	ML grade
x1	100
x2	80
...	...

- Possible solution (nearest neighbor classifier):

1. For any student x find the “closest" student x_i in the training set
2. Predict y_i, the grade of the closest student

Slide credit: Tommi Jaakkola, MIT

Evaluation

- How can we tell how good our predictions are?
  - we can wait till the end of this course...
  - we can try to assess the accuracy based on the data we already have (training data)
- Possible solution:
  - divide the training set further into training and validation sets;
  - evaluate the classifier constructed on the basis of only the smaller training set on the new validation set

Slide credit: Tommi Jaakkola, MIT

Model Selection

- We can refine
  - the estimation algorithm (e.g., using a classifier other than the nearest neighbor classier)
  - the representation (e.g., base the summaries on a different set of courses)
  - the assumptions (e.g., perhaps students work in groups) etc.
- We have to rely on the method of evaluating the accuracy of our predictions to select among the possible refinements

Slide credit: Tommi Jaakkola, MIT