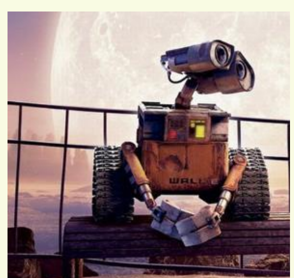
Why do we need Machine Learning?

- ML is a very exciting field !!!
- It’s used in applications that cannot be explicitly programmed.

OCR

→ Computer Program → Thank you

Robot Navigation



Work well with learning algorithms

Why do we need Machine Learning?

- ML has lots of applications in many areas:
  - Media
  - Biology
  - Data Bases
  - Economics
  - Internet/Social Networks
  - Robots
  - Games
  - …

What is Learning?

- Learning is an essential human property
- Learning: Acquisition of knowledge, understanding, and ability with experience.
- Learning IS NOT learning by heart
- Any computer can learn by heart, the difficulty is to make a prediction – generalize a behavior to a novel situation.

Machine Learning

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

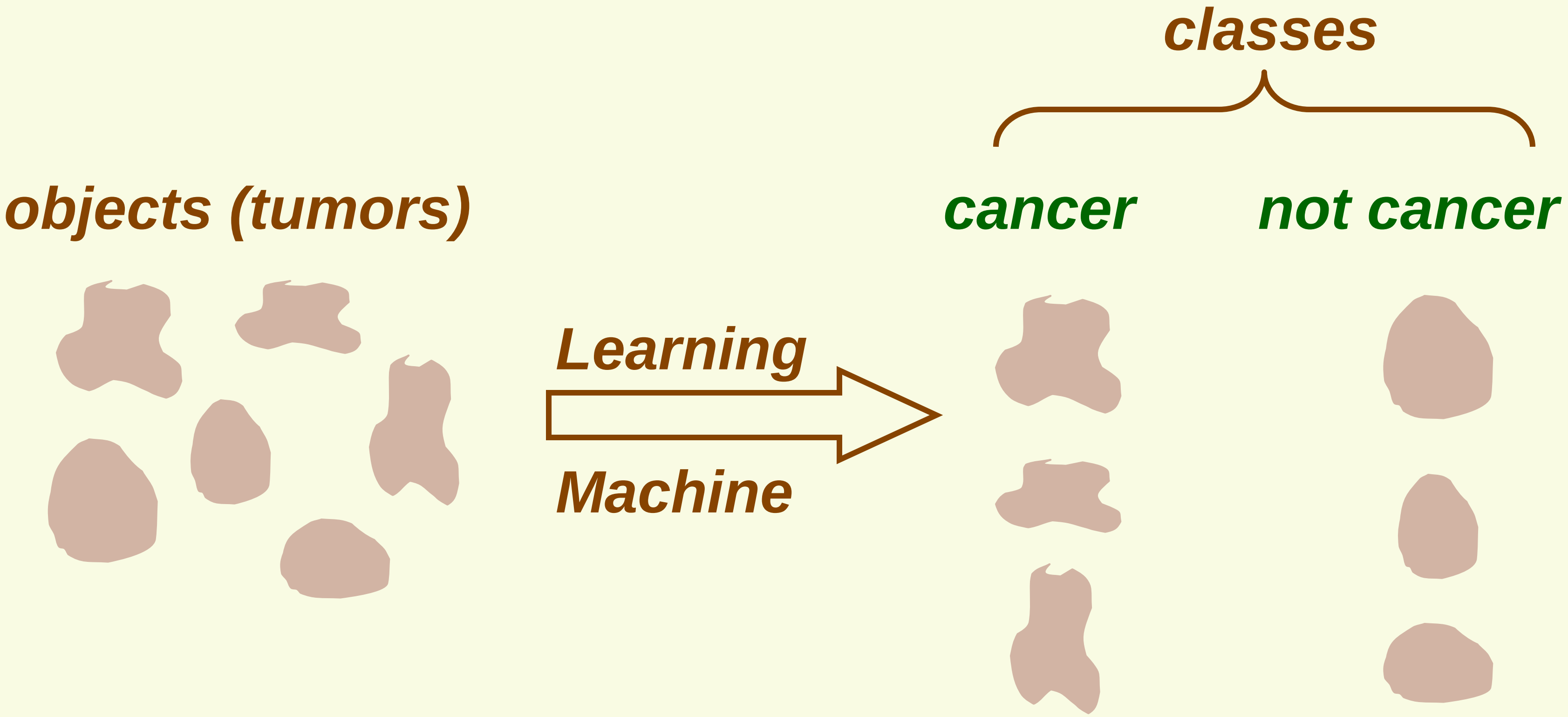
OCR

Experience: images of a handwritten “a”.

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

Measure of Performance: recognition rate.

Application: Medical diagnostics



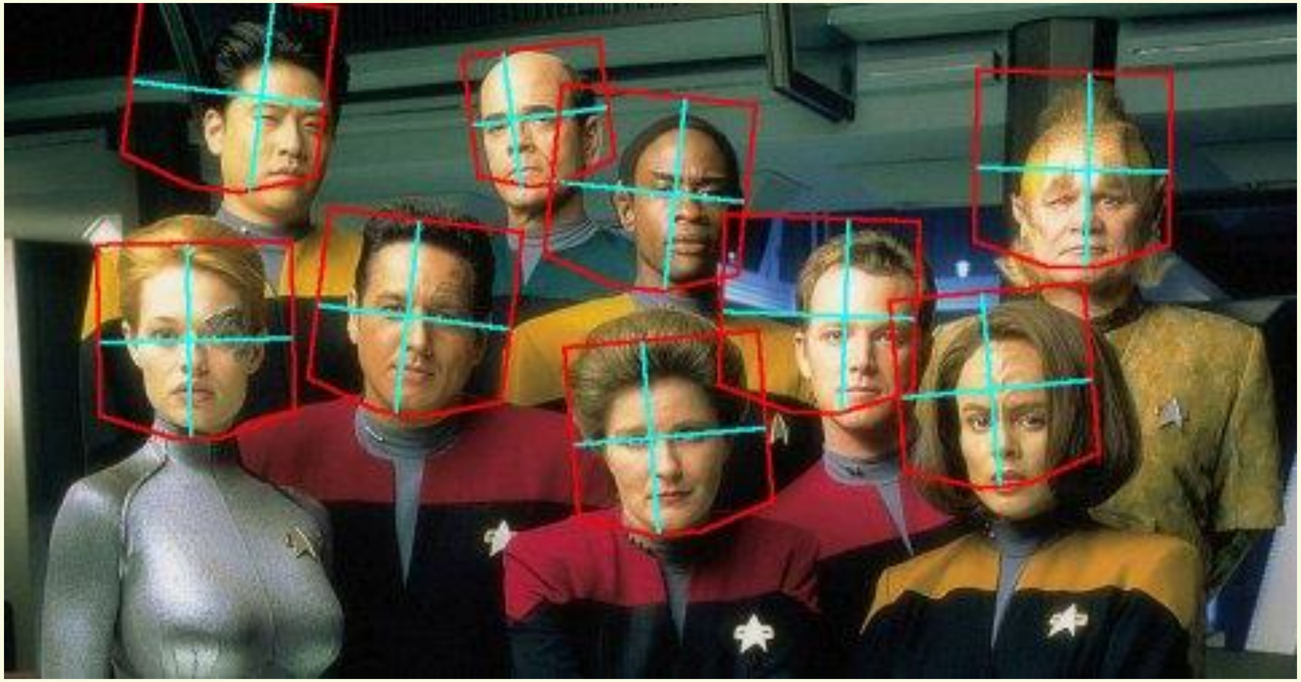
Application: Loan applications

classes

objects (people)

	income	debt	married	age	approve	deny
John Smith	200,000	0	yes	80		☑
Peter White	60,000	1,000	no	30	☑	
Ann Clark	100,000	10,000	yes	40	☑	
Susan Ho	0	20,000	no	25		☑

Application: Face Detection



Application: Text Classification

→ Company homepage vs. Personal homepage

Other Applications

- speech recognition, speaker recognition/verification
- security: face recognition, event detection in videos
- adaptive control: navigation of mobile robots…
- fraud detection: e.g. detection of “unusual” usage patterns for credit cards or calling cards
- spam filtering
- financial prediction (many people on Wall Street use machine learning)
- Many others

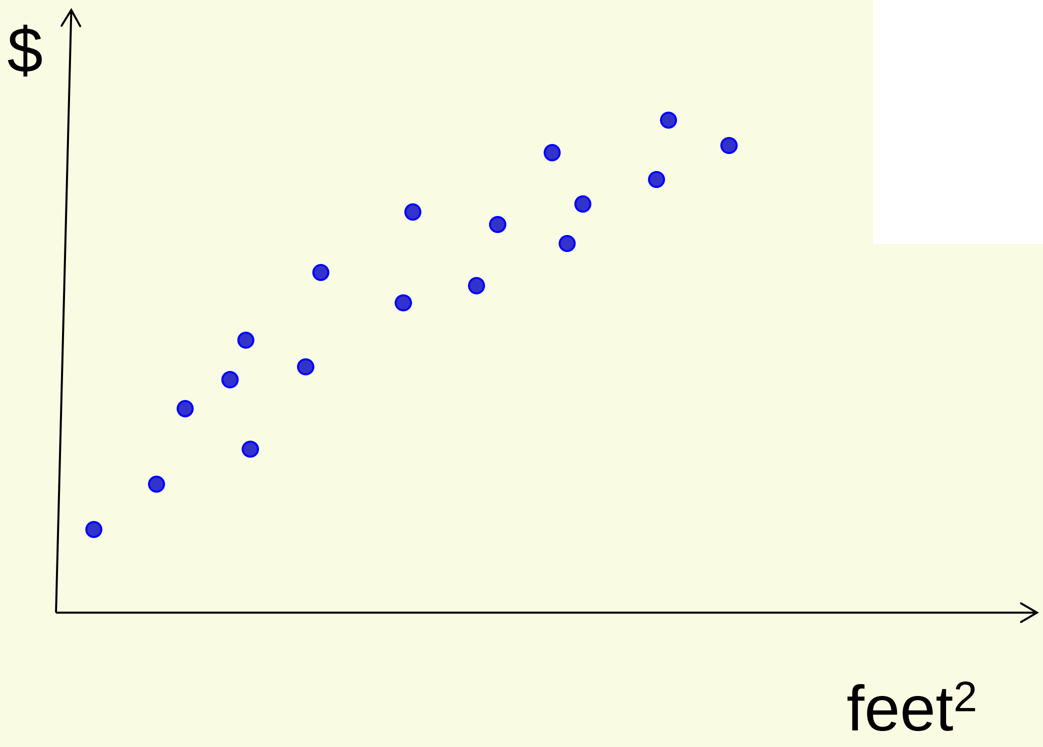
Types of Learning Problems

- Supervised learning: given a set of training inputs and corresponding outputs (correct answers), produce the "correct" outputs for the new inputs.
  - classification, regression

Regression

- Housing Prices

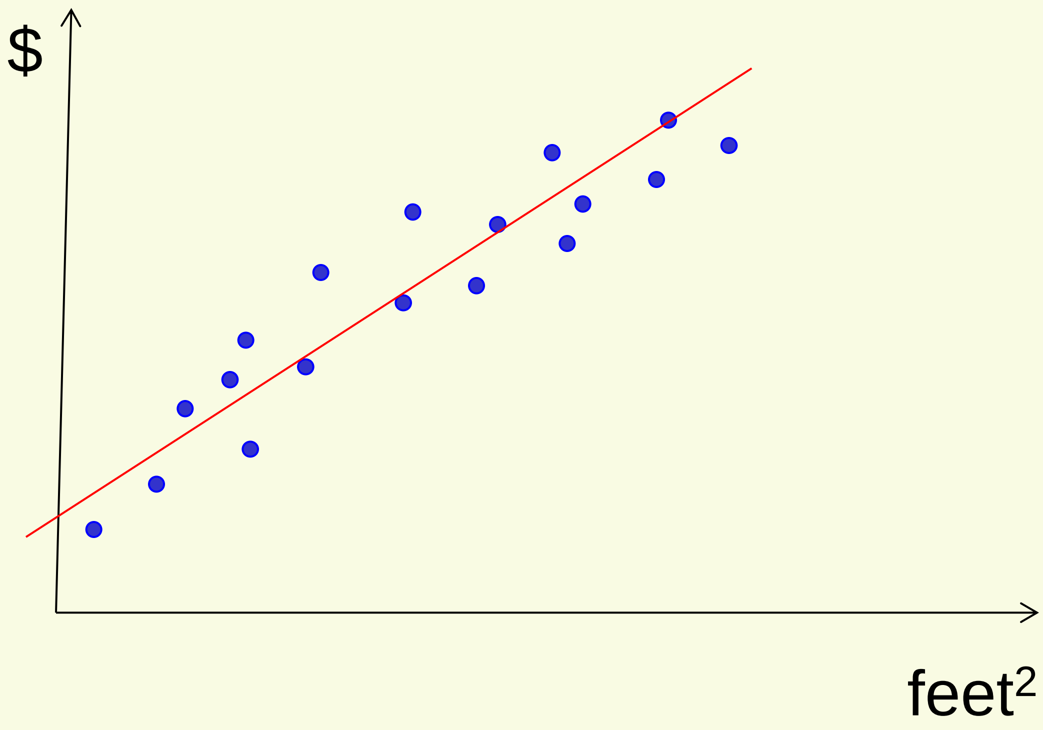
Living area ($feet^2$)	Price (1000\$s)
2104	400
1600	330
2400	369
1416	232
3000	540
⋮	⋮



Example by Andrew Ng, Stanford

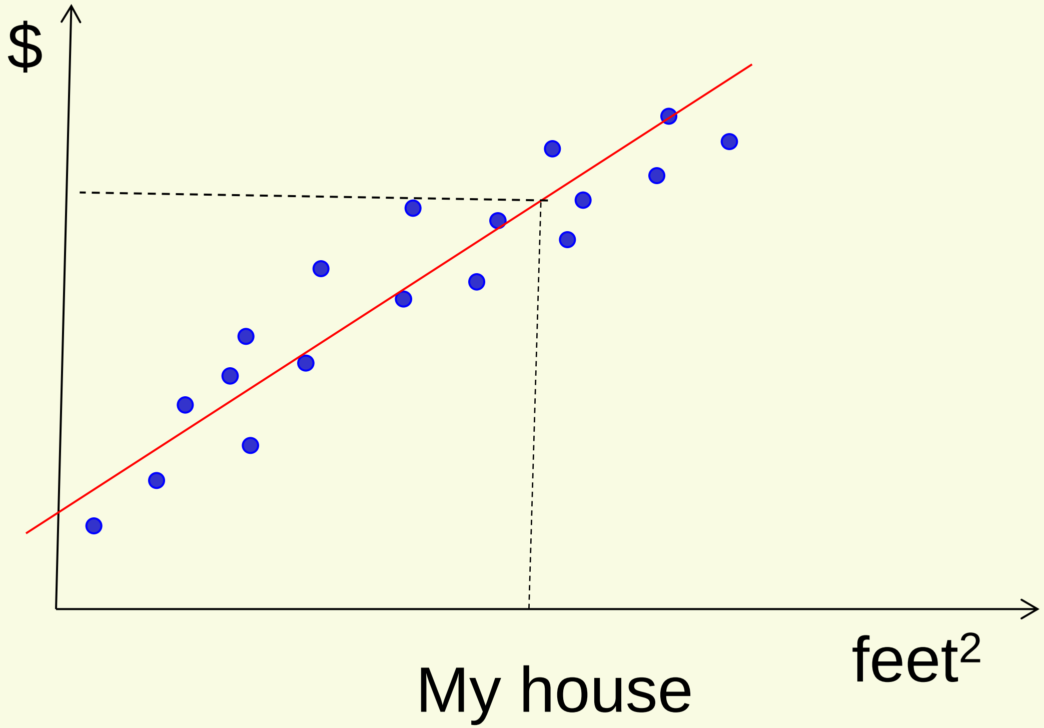
Regression

- Housing Prices



Regression

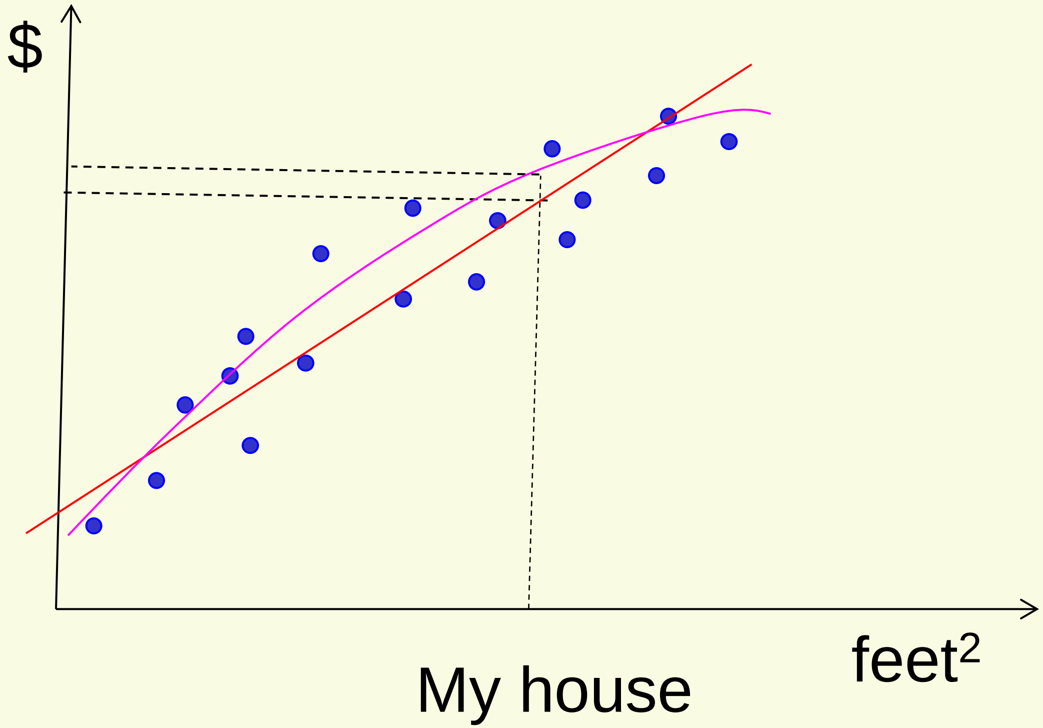
- Housing Prices



Example by Andrew Ng, Stanford

Regression

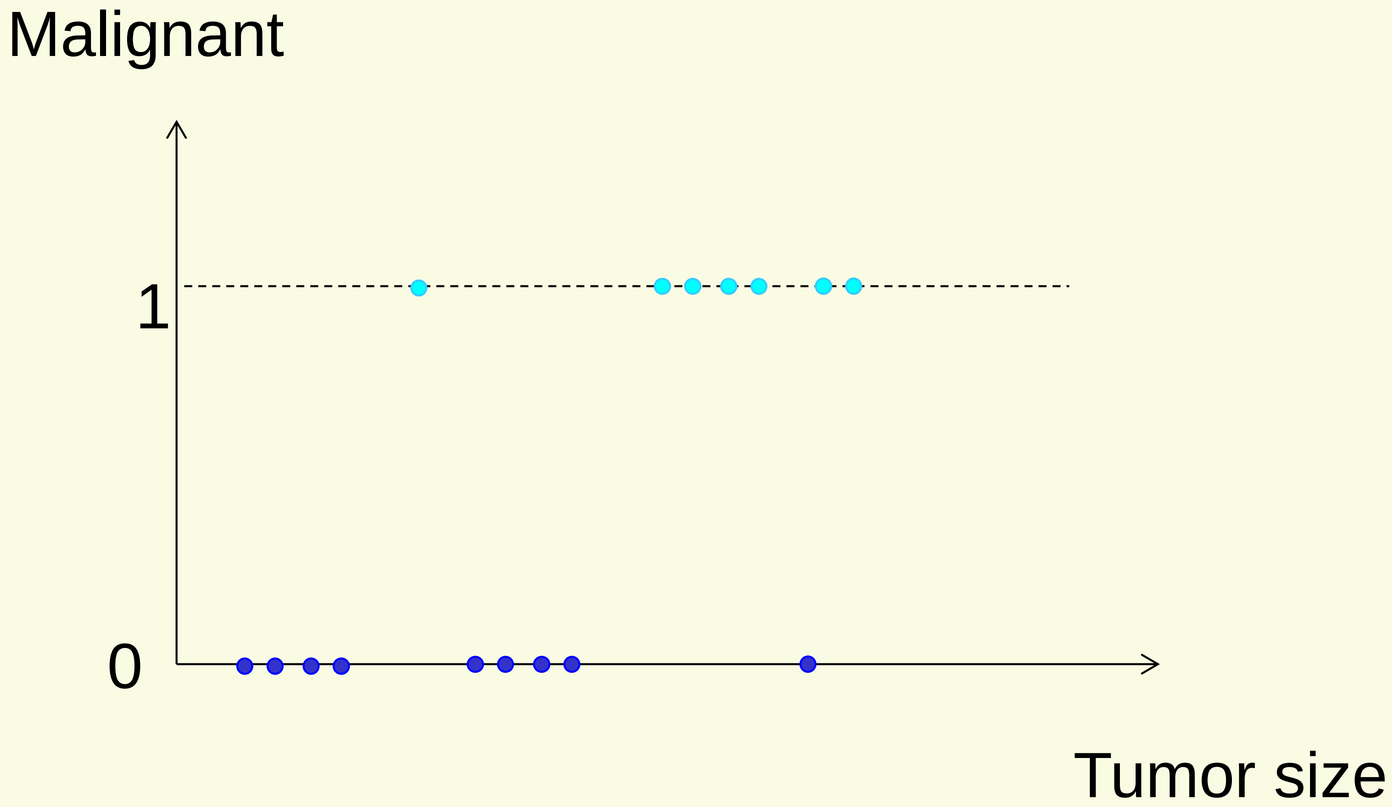
- Housing Prices



Example by Andrew Ng, Stanford

Classification

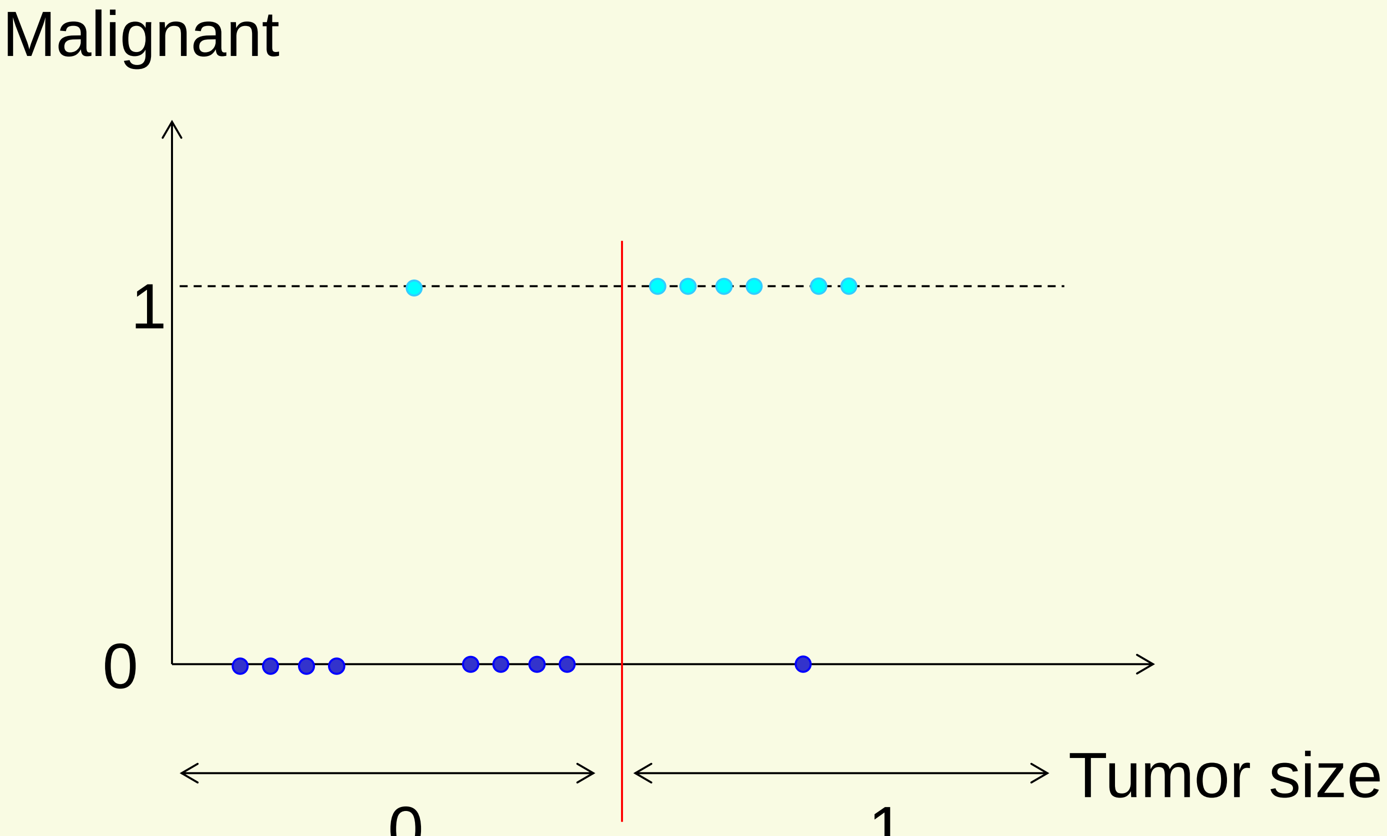
- Tumor Classification



Example by Andrew Ng, Stanford

Classification

- Tumor Classification

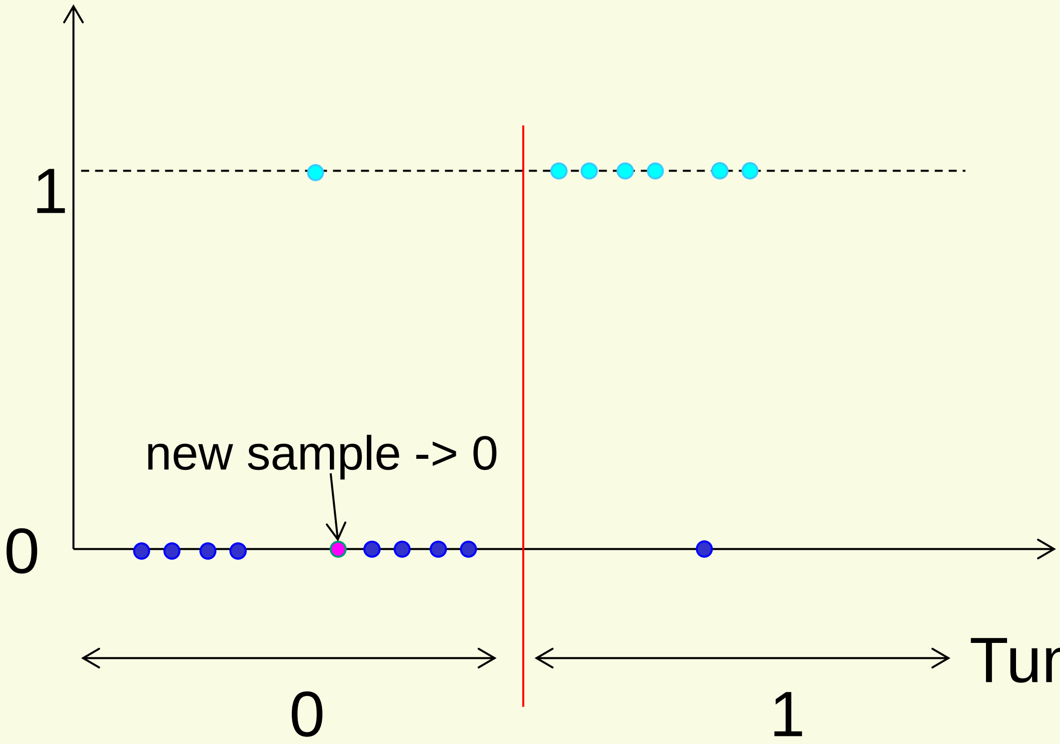


Example by Andrew Ng, Stanford

Classification

- Tumor Classification

Malignant

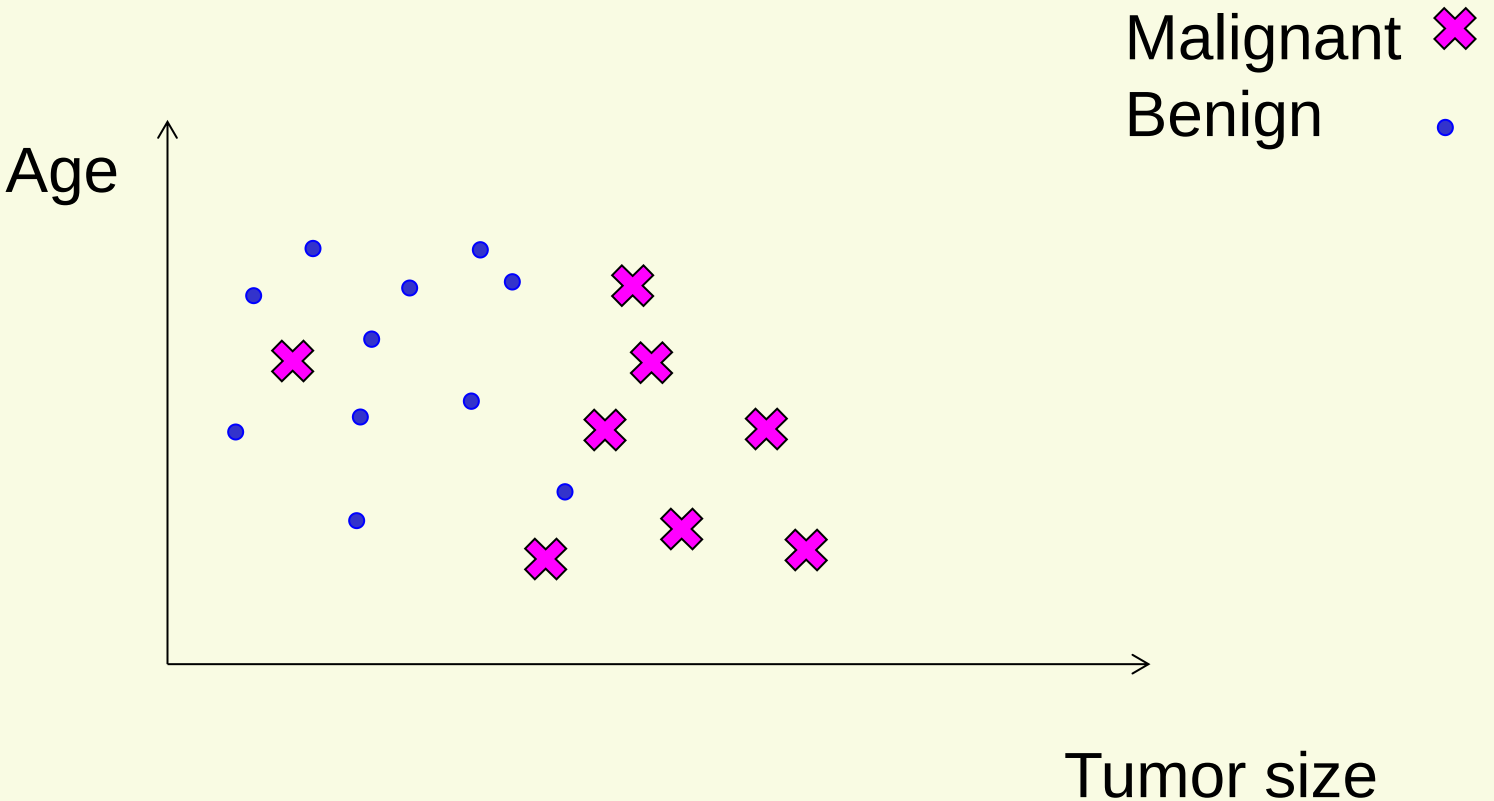


Tumor size

Example by Andrew Ng, Stanford

Classification

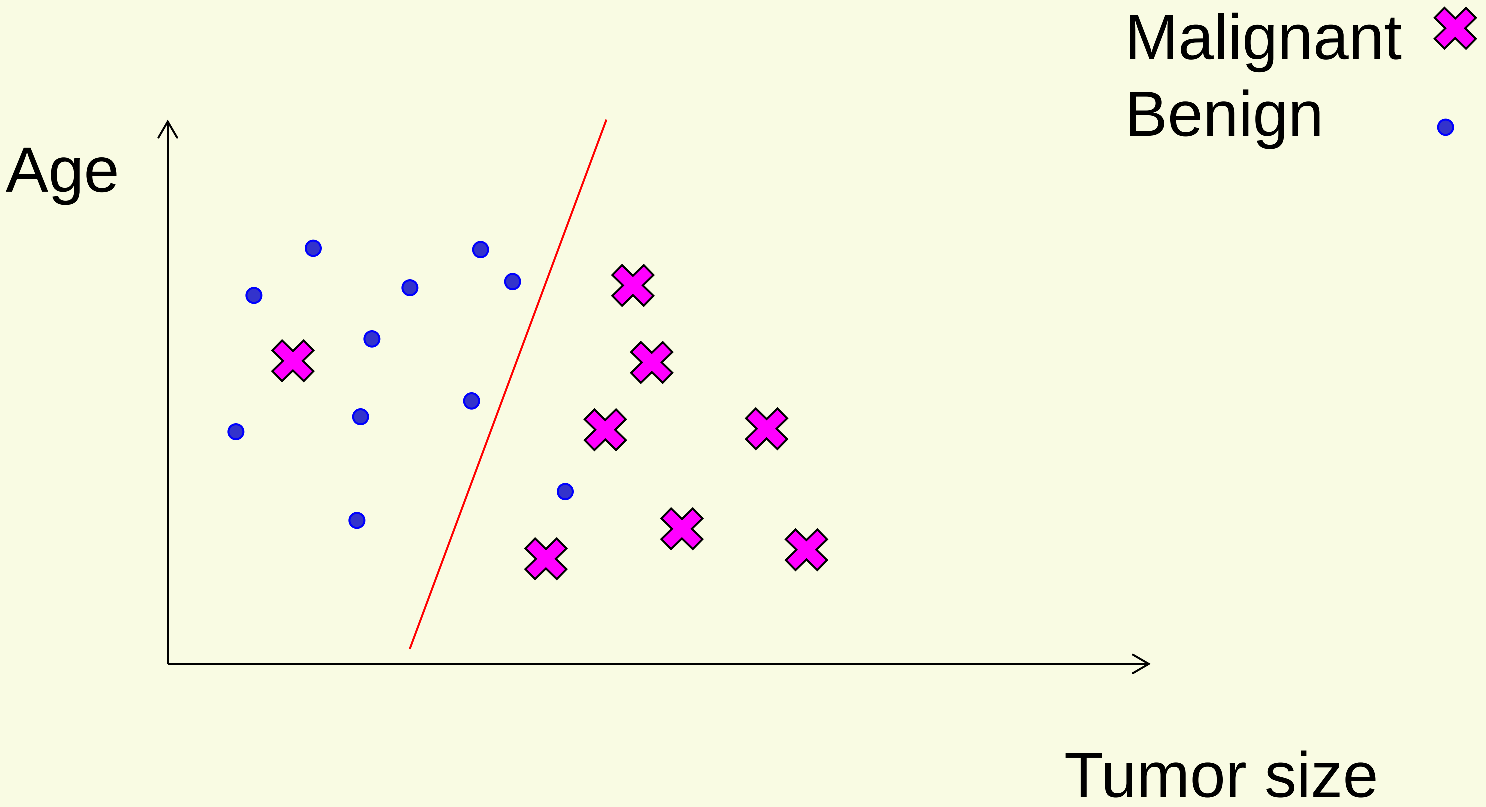
- Tumor Classification



Example by Andrew Ng, Stanford

Classification

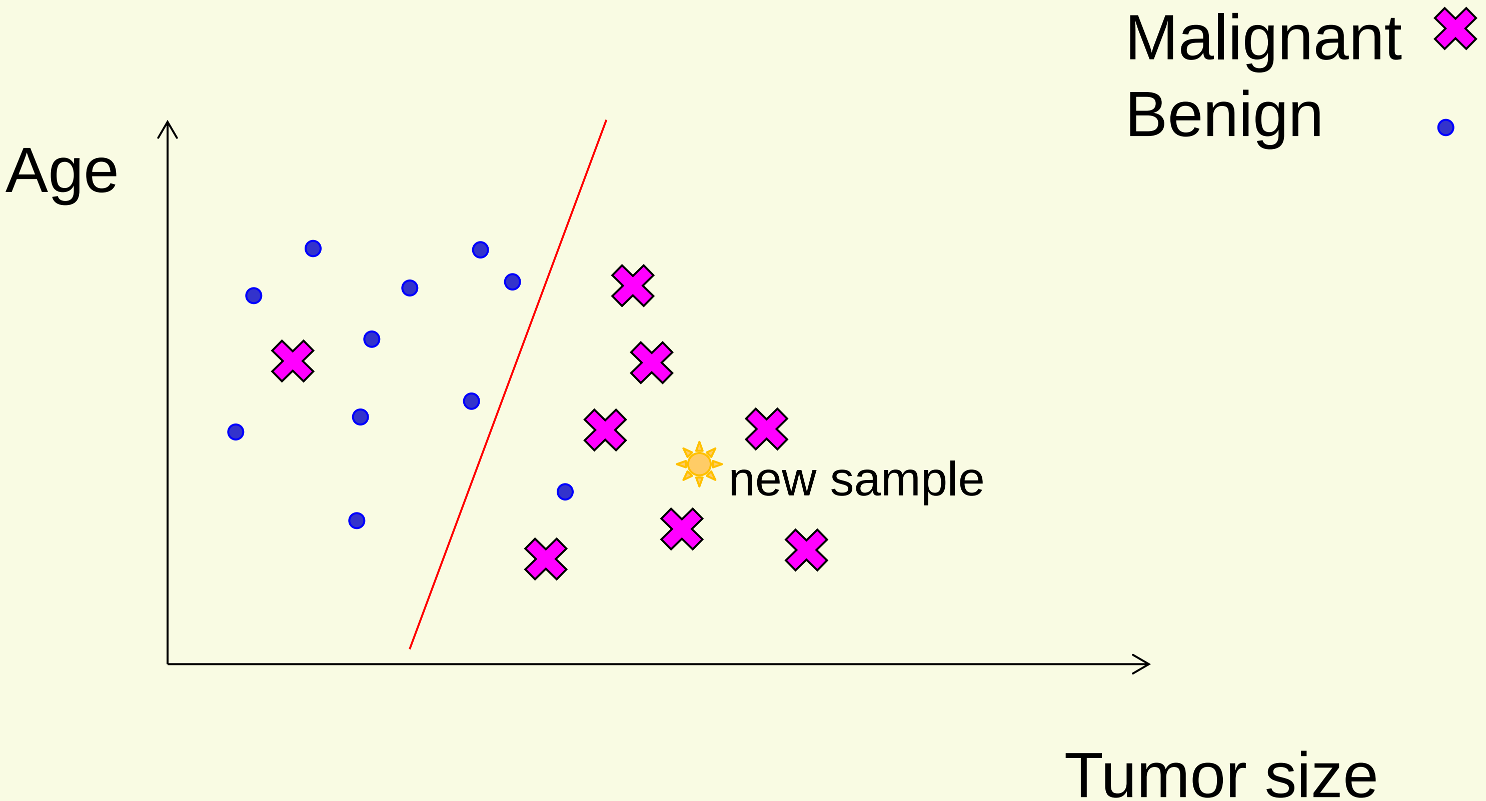
- Tumor Classification



Example by Andrew Ng, Stanford

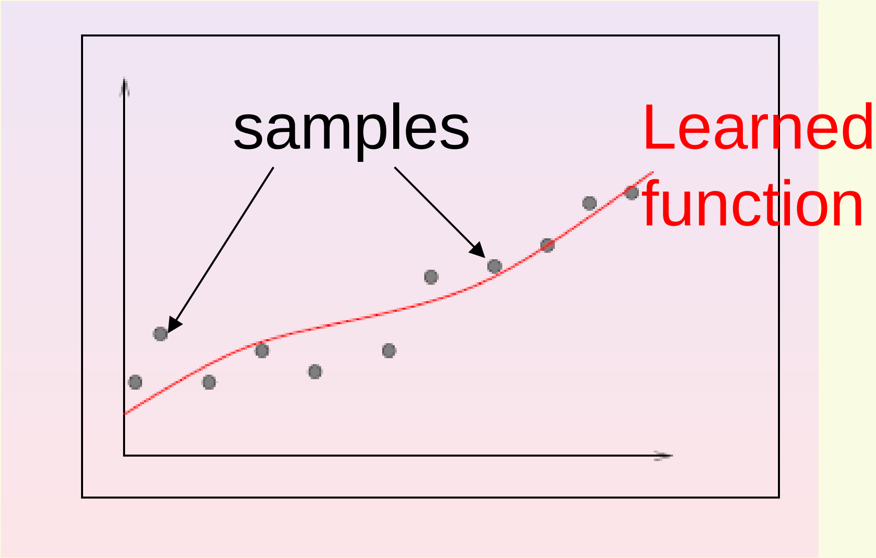
Classification

- Tumor Classification

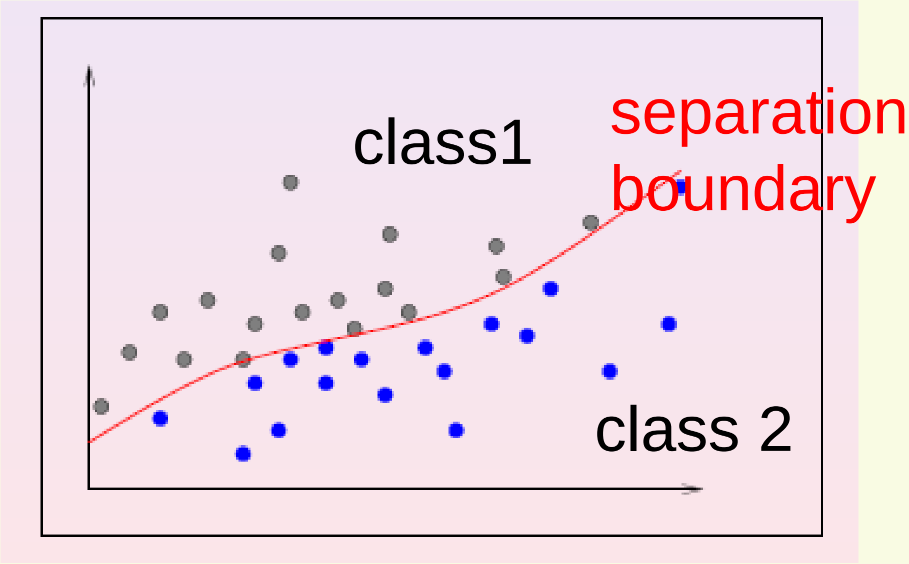


Example by Andrew Ng, Stanford

Two kinds of Supervised Learning



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

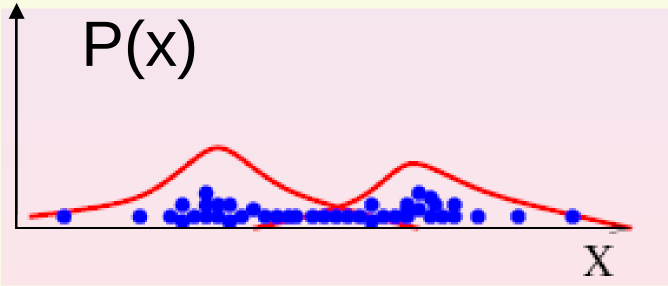


- Classification: outputs are discrete variables (category labels). Learn a decision boundary that separates one class from the other.

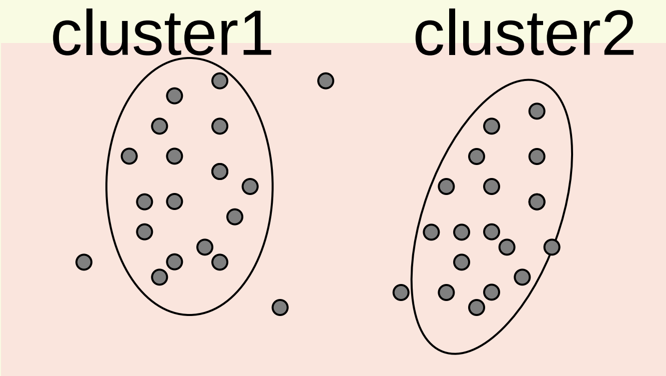
Types of Learning Problems

- Supervised learning: given a set of training inputs and corresponding outputs, produce the “correct” outputs for new inputs.
  - classification, regression
- 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
  - clustering, density estimation, embedding

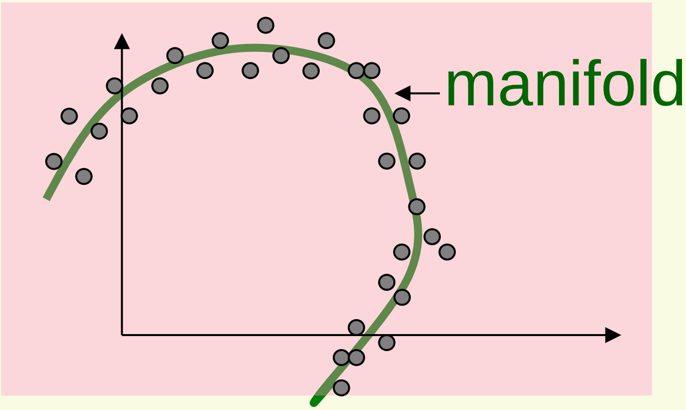
Unsupervised Learning



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



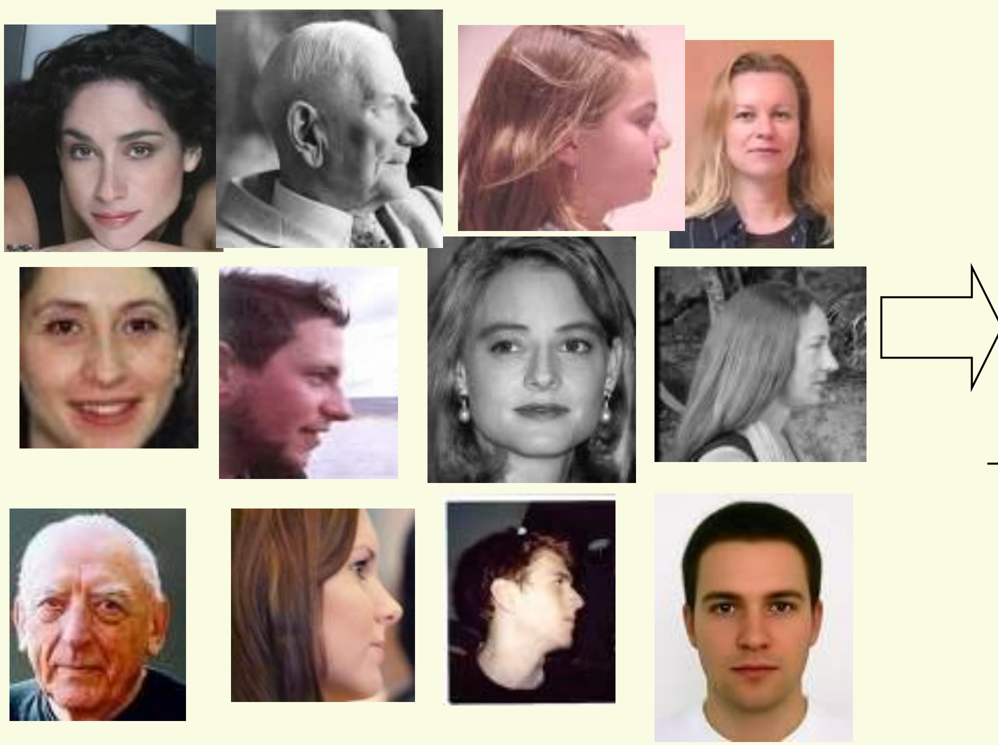
- Clustering: discover "clumps" of points

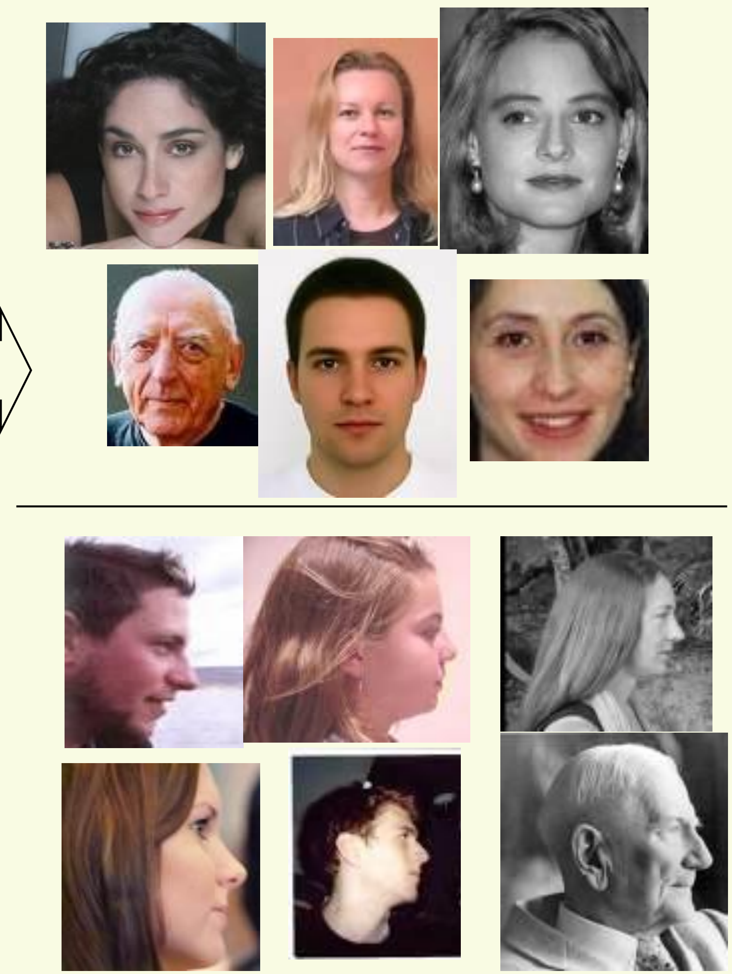


- Embedding: discover low-dimensional manifold or surface near which the data lives.

Clustering Example

- Cluster images of faces into two groups

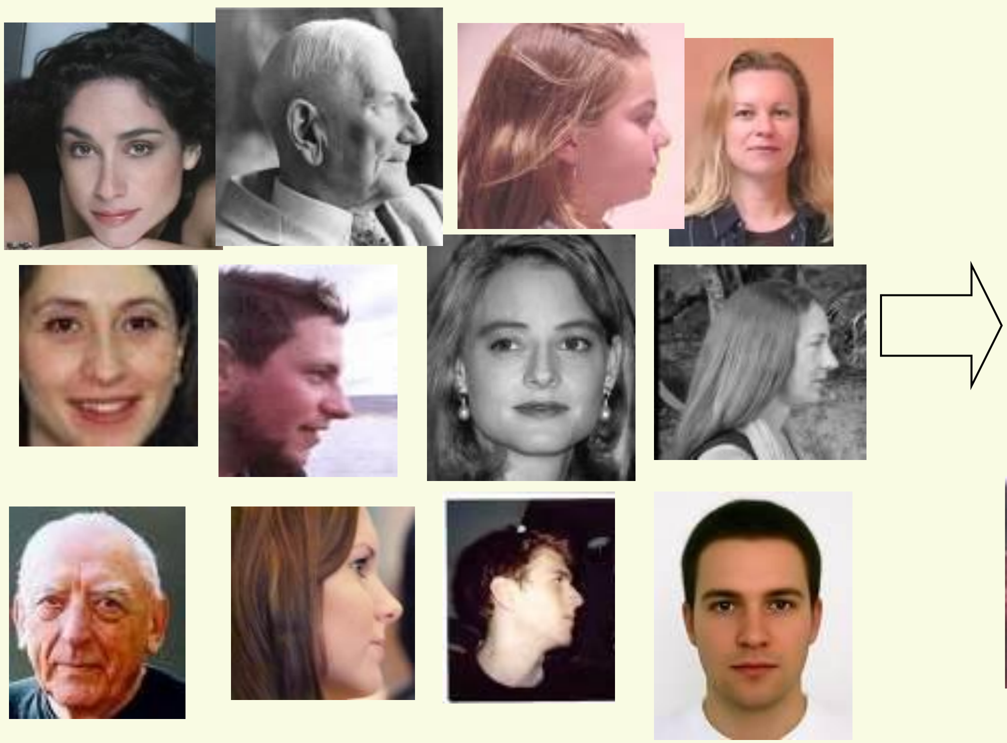


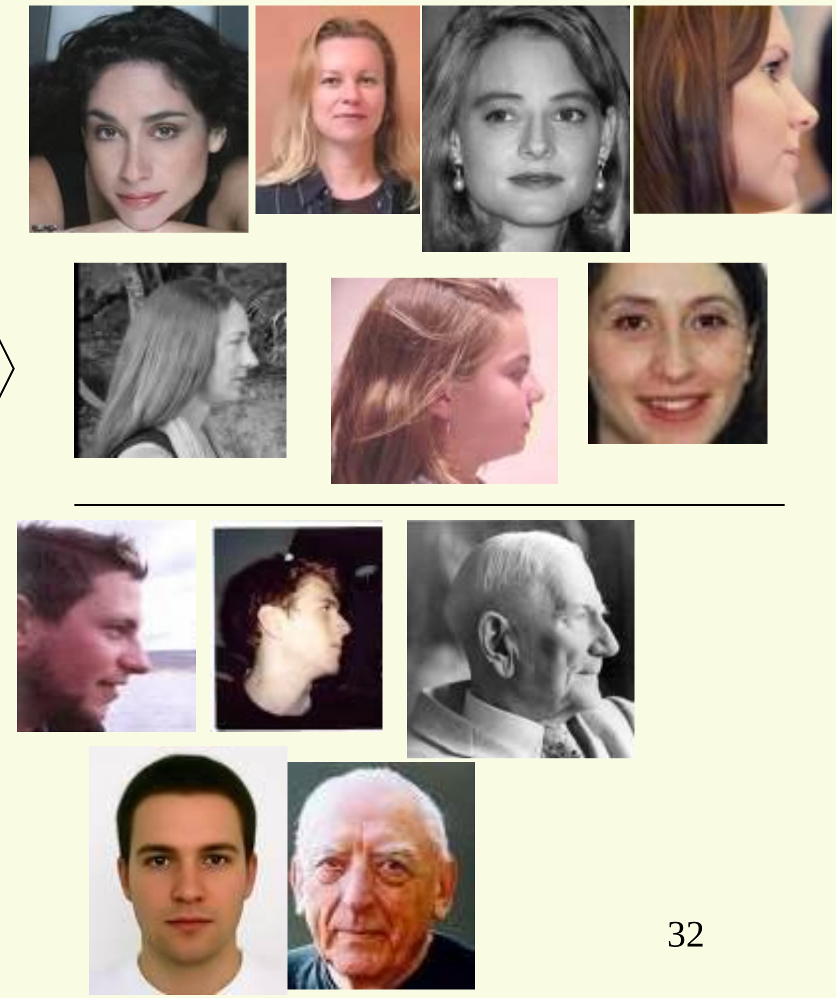


by view point

Clustering Example

- Cluster images of faces into two groups





by gender

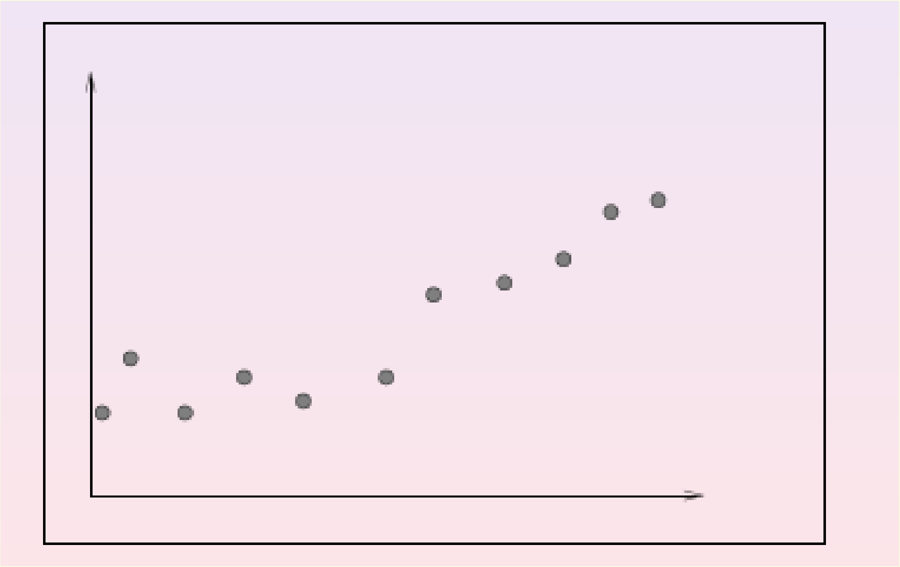
Types of Learning Problems

- Supervised learning: given a set of training inputs and corresponding outputs, produce the “correct” outputs for new inputs.
  - classification, regression
- 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
  - clustering, density estimation, embedding
- Reinforcement learning, where we only get feedback in the form of how well we are doing (For example the outcome of the game).

  I won’t talk much about that in this course.
  - planning

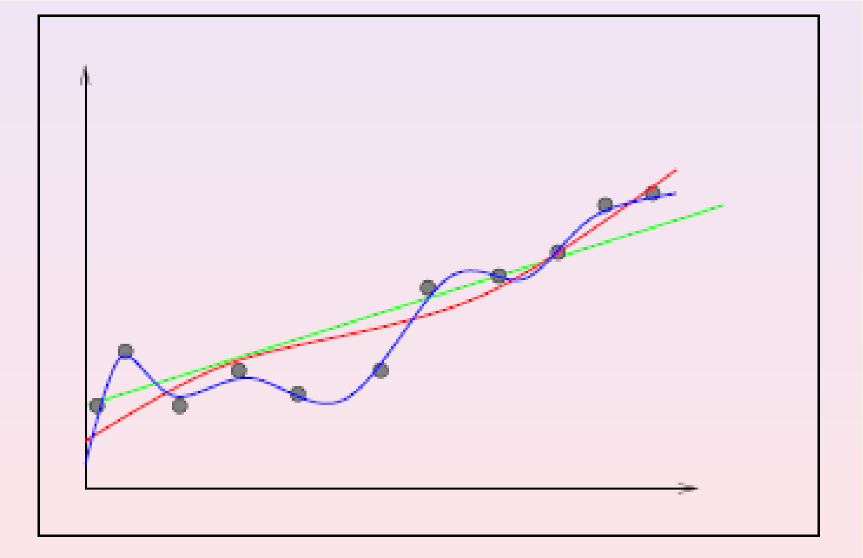
Why Learning is Difficult?

- 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?

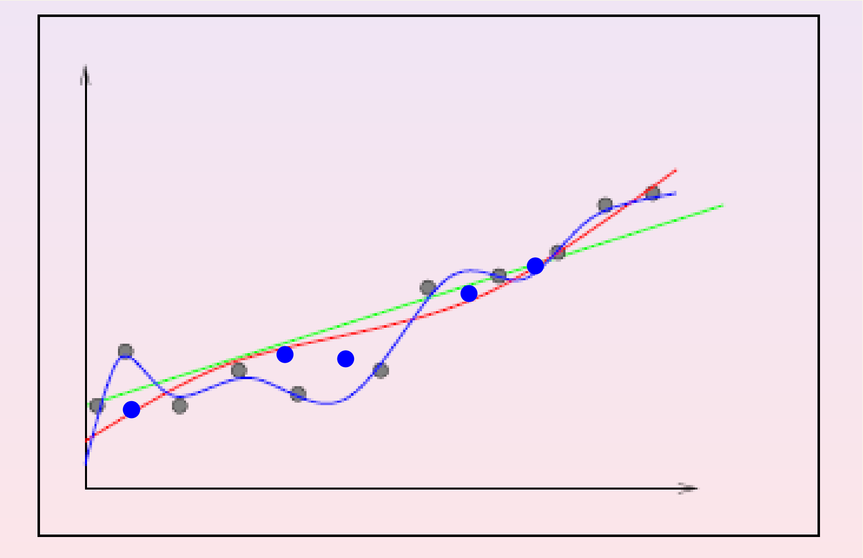
- Given a finite amount of training data, you have to derive a relation for an infinite domain
- In fact, there is an infinite number of such relations



- 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):

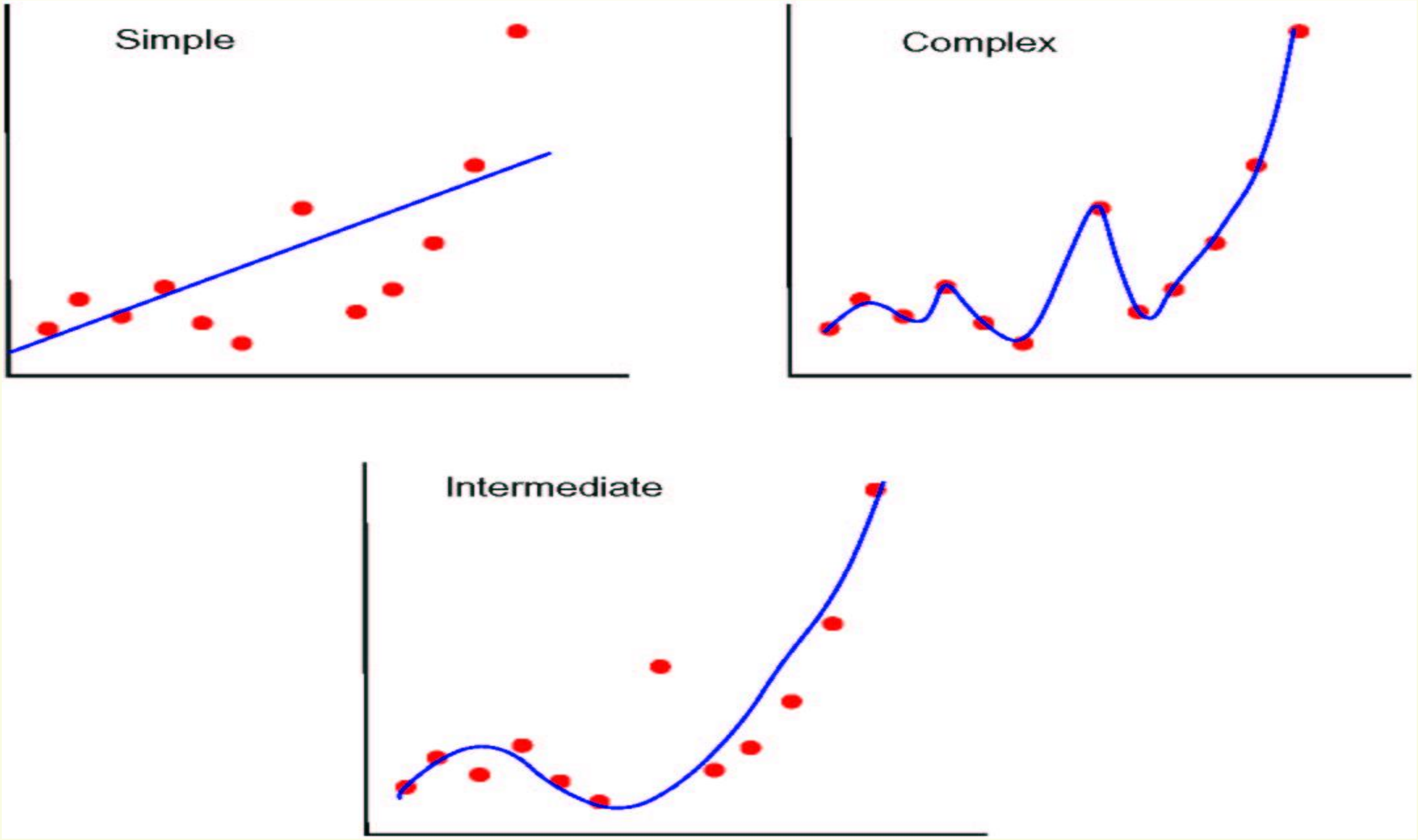
  One should not increase, beyond what is necessary,

  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?

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

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

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

- 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

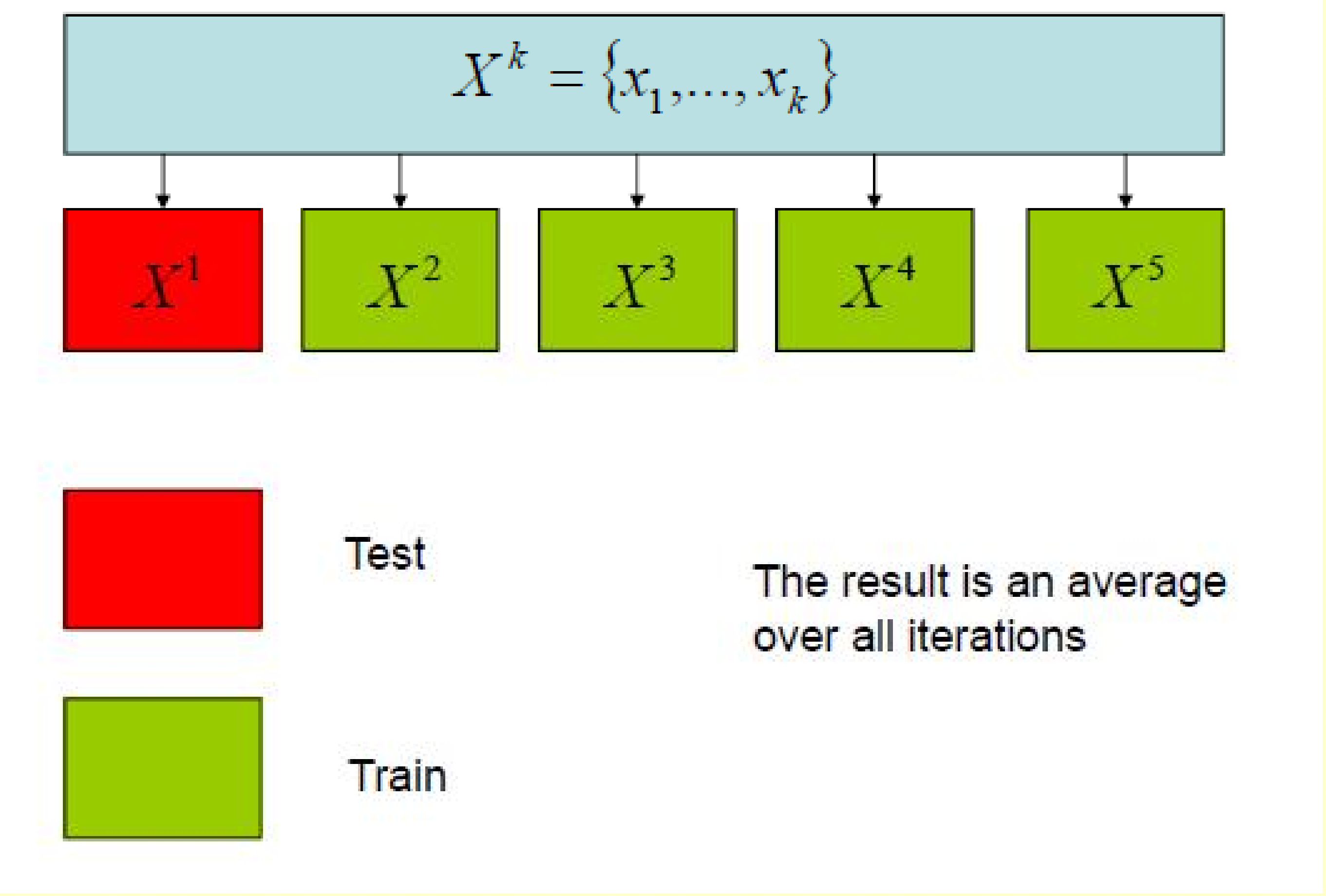
Model Selection

- 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

Validation of Model

- Cross Validation



Interpreting Results

=== Confusion Matrix ===

PREDICTED CLASS

a	b	<-- classified as	
TP	FN	a = Stress	ACTUAL CLASS
FP	TN	b = Baseline	

Example:

=== Confusion Matrix ===

PREDICTED CLASS

a	b	<-- classified as	
266	63	a = Stress	ACTUAL CLASS
57	214	b = Baseline	

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Sensitivity(Recall)} = \frac{TP}{TP + FN} \qquad \text{Specificity} = \frac{TN}{TN + FP}$$

Over-fitting