Machine learning: lecture 1

Tommi S. Jaakkola
MIT CSAIL
tommi@csail.mit.edu

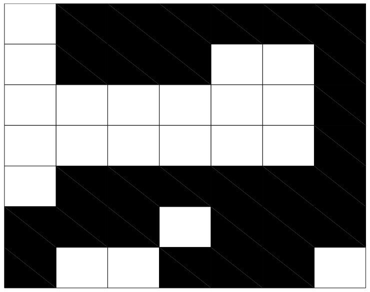
6.867 Machine learning: administrivia

- Instructor: Prof. Tommi Jaakkola (tommi@csail.mit.edu)
- TA: Jason Johnson (jasonj@mit.edu)
- General info
  - lectures MW 2.30-4pm in 35-225
  - tutorials/recitations (time/location tba)
  - website `http://www.ai.mit.edu/courses/6.867`
- Grading
  - midterm (15%), final (25%)
  - 6 ($\approx$ bi-weekly) problem sets (30%)
  - final project (30%)

Machine learning

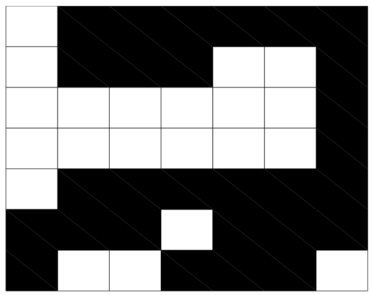
- Statistical machine learning
  - principles, methods, and algorithms for learning and prediction on the basis of past experience
  - already everywhere: speech recognition, hand-written character recognition, information retrieval, operating systems, compilers, fraud detection, security, defense applications, ...

Learning

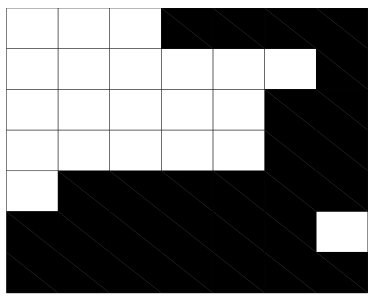


"yes"

Learning

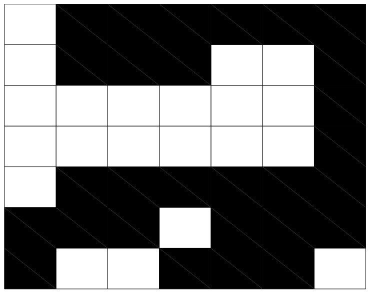


“yes”

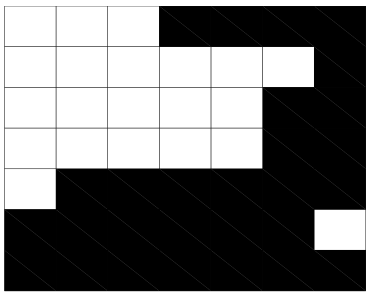


“yes”

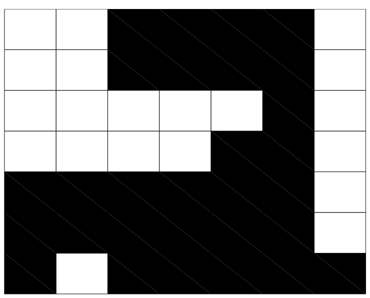
Learning



"yes"



"yes"



"no" (oops)

Learning

- Steps
  - entertain a (biased) set of possibilities
  - adjust predictions based on feedback
  - rethink the set of possibilities

Learning

- Steps
  - entertain a (biased) set of possibilities
  - adjust predictions based on feedback
  - rethink the set of possibilities
- Principles of learning are "universal"
  - society (e.g., scientific community)
  - animal (e.g., human)
  - machine

Learning and prediction

- We make predictions all the time but rarely investigate the processes underlying our predictions
- In carrying out scientific research we are also governed by how theories are evaluated
- To automate the process of making predictions we need to understand *in addition* how we search and refine "theories"

Learning: key steps

- data and assumptions
  - what data is available for the learning task?
  - what can we assume about the problem?

Learning: key steps

- data and assumptions
  - what data is available for the learning task?
  - what can we assume about the problem?
- representation
  - how should we represent the examples to be classified

Learning: key steps

- data and assumptions
  - what data is available for the learning task?
  - what can we assume about the problem?
- representation
  - how should we represent the examples to be classified
- method and estimation
  - what are the possible hypotheses?
  - how do we adjust our predictions based on the feedback?

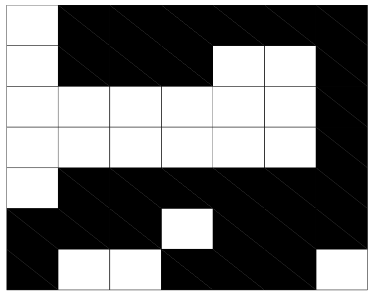
Learning: key steps

- data and assumptions
  - what data is available for the learning task?
  - what can we assume about the problem?
- representation
  - how should we represent the examples to be classified
- method and estimation
  - what are the possible hypotheses?
  - how do we adjust our predictions based on the feedback?
- evaluation
  - how well are we doing?

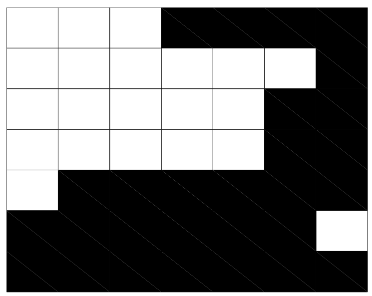
Learning: key steps

- data and assumptions
  - what data is available for the learning task?
  - what can we assume about the problem?
- representation
  - how should we represent the examples to be classified
- method and estimation
  - what are the possible hypotheses?
  - how do we adjust our predictions based on the feedback?
- evaluation
  - how well are we doing?
- model selection
  - can we rethink the approach to do even better?

Data and assumptions



"yes"



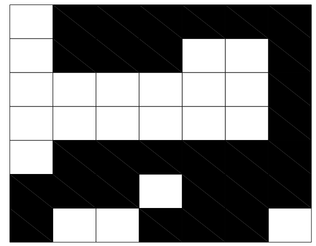
"yes"

. . .   . . .

– is this a digit/character/image classification problem?
– how are the examples generated/labeled?

Representation

- There are many ways of presenting the same information



01111110011100100000001000000010011111101110111110011101111110001

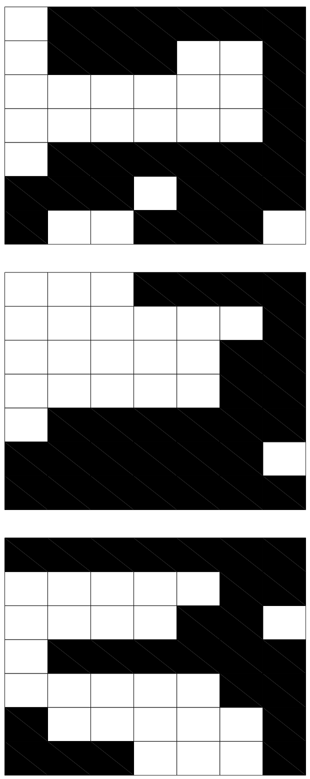
- The choice of representation may determine whether the learning task is very easy or very difficult

Representation

0111111001110010000000100000001001111110111011111001110111110001	“yes”
0001111100000011000001110000011001111110111111001111111100000011	“yes”
1111111000000110000011000111111000000111100000111110001101111111	“no”

Representation

01111110011100100000000100000000100111111011101111100111011111110001	“yes”
00011111000000110000011100000110011111101111110011111111100000011	“yes”
11111110000001100000110001111110000001111000001111100011011111111	“no”



“yes”

“yes”

“no”

Method and estimation

- Examples (binary vectors of length $d = 64$)

$$\mathbf{x} = [111111100\ldots0000110001101111111]^T$$

- Labels $y \in \{-1, 1\}$ ("no","yes")
- A linear classifier (hypotheses)

$$\hat{y} = \text{sign}\,(\,\theta \cdot \mathbf{x}\,) = \text{sign}\left(\sum_{i=1}^{d} \theta_i x_i\right)$$

where θ is a vector of *parameters* we have to learn.

Method and estimation cont'd

$\mathbf{x}$	y
0111111001110010000000100000001001111110111011111001110111110001	+1
0001111100000011000001110000011001111110111111001111111100000011	+1
1111111000000110000011000111111000000111100000111110001101111111	−1
. . .	. . .

- How do we adjust the parameters θ based on the labels?

$$\hat{y} = \operatorname{sign}\left(\, \theta \cdot \mathbf{x} \,\right)$$

Method and estimation cont'd

$\mathbf{x}$	y
0111111001110010000000010000000100111111011101111110011101111110001	+1
00011111000000011000000111000000110011111101111110011111111100000011	+1
11111110000000110000001100011111100000001111000000111110001101111111	−1
. . .	. . .

- How do we adjust the parameters θ based on the labels?

$$\hat{y} = \text{sign}\,(\,\theta \cdot \mathbf{x}\,)$$

For example, we can simply refine/update the parameters whenever we make a mistake

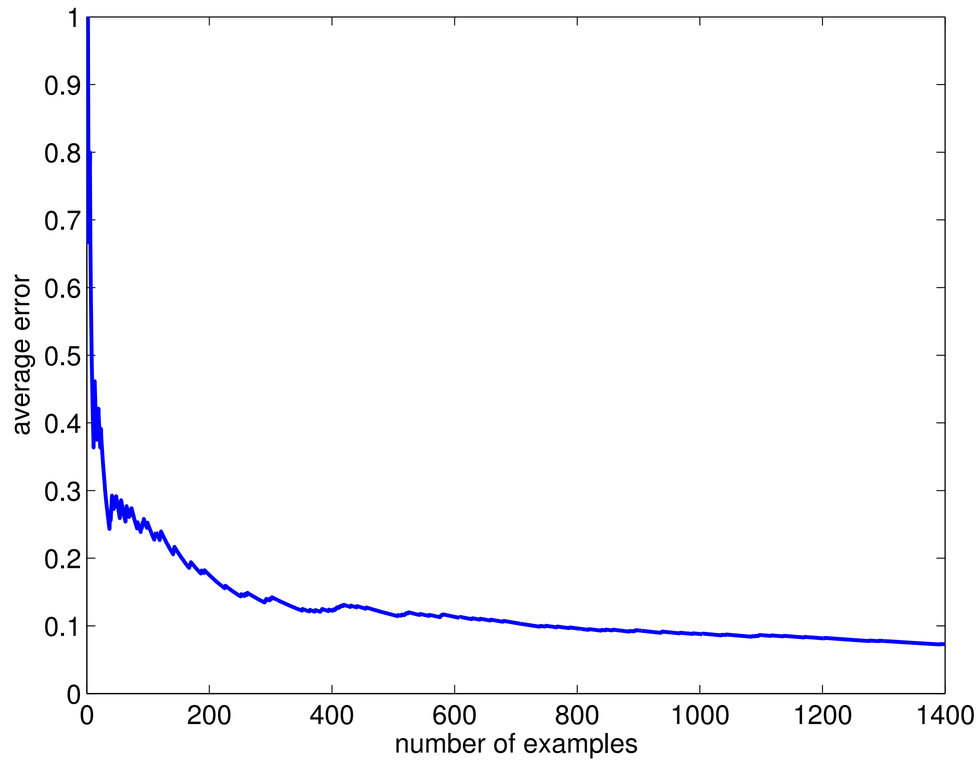
$$\theta \leftarrow \theta + y\mathbf{x} \;\; \text{when prediction was wrong}$$

Evaluation

- How do we measure how well the method is working?

Evaluation

- How do we measure how well the method is working?



For example: average classification/prediction error as a function of the number of examples seen so far

Model selection

- Can we rethink the approach to do even better?
  - our classifier is limited, can we make it more flexible?
  - is there an entirely different type of classifier that would be more suitable?

Types of learning problems

A rough classification of learning problems:

- Supervised learning
  - explicit feedback (e.g., labels)
- Unsupervised learning
  - no feedback, emphasis on structure and organization
- Semi-supervised learning
  - partial feedback (e.g., a few labels, mostly unlabeled)
- Reinforcement learning
  - delayed feedback