Machine learning: lecture 1

Tommi S. Jaakkola
MIT AI Lab
tommi@ai.mit.edu

6.867 Machine learning: administrivia

- Instructor: Prof. Tommi Jaakkola (tommi@ai.mit.edu)
- TA: Gregory Shakhnarovich (gregory@ai.mit.edu)

- General info
  - lectures TR 2.30-4pm in 37-212
  - tutorials/recitations (time/location tba)
  - website `http://www.ai.mit.edu/courses/6.867` (please register)

- Grading
  - midterm (15%), final (25%)
  - 6 ($\approx$ bi-weekly) problem sets (30%)
  - final project (30%)

Broader context

- What is learning anyway?

- Principles of learning are “universal”
  - society (e.g., scientific community)
  - animal (e.g., human)
  - machine

- Prediction is the key...

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"

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, ...

Example

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

Example

- A classifcation problem: predict the grades for students taking this course
- Key steps:
  1. data
  2. assumptions
  3. representation
  4. estimation
  5. evaluation
  6. model selection

Example

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

- Key steps:

1. data: what "past experience" can we rely on?

Example

- A classifcation problem: predict the grades for students taking this course
- Key steps:
  1. data: what "past experience" can we rely on?
  2. assumptions: what can we assume about the students or the course?

Example

- A classifcation problem: predict the grades for students taking this course
- Key steps:
  1. data: what "past experience" can we rely on?
  2. assumptions: what can we assume about the students or the course?
  3. representation: how do we "summarize" a student?

Example

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

- Key steps:
  1. data: what “past experience” can we rely on?
  2. assumptions: what can we assume about the students or the course?
  3. representation: how do we “summarize” a student?
  4. estimation: how do we construct a map from students to grades?

Example

- A classifcation problem: predict the grades for students taking this course
- Key steps:
  1. data: what "past experience" can we rely on?
  2. assumptions: what can we assume about the students or the course?
  3. representation: how do we "summarize" a student?
  4. estimation: how do we construct a map from students to grades?
  5. evaluation: how well are we predicting?

Example

- A classifcation problem: predict the grades for students taking this course
- Key steps:
  1. data: what "past experience" can we rely on?
  2. assumptions: what can we assume about the students or the course?
  3. representation: how do we "summarize" a student?
  4. estimation: how do we construct a map from students to grades?
  5. evaluation: how well are we predicting?
  6. model selection: perhaps we can do even better?

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

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	course 1	course 2	...
Peter	A	B	A	...
David	B	A	A	...

- "test" data:

Student	ML	course 1	course 2	...
Jack	?	C	A	...
Kate	?	A	A	...

- Anything else we could use?

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

Representation

- 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

$$\mathbf{x}_i = [A\ C\ B]$$

  where the grades correspond to (say) 18.06, 6.041, and 6.034.
- The available data in this representation

Training		Test	
Student	ML grade	Student	ML grade
$\mathbf{x}_1$	A	$\mathbf{x}'_1$	?
$\mathbf{x}_2$	B	$\mathbf{x}'_2$	?
...	...	...	...

Estimation

- Given the training data

Student	ML grade
$\mathbf{x}_1$	A
$\mathbf{x}_2$	B
...	...

  we need to find a mapping from "input vectors" $\mathbf{x}$ to "labels" y encoding the grades for the ML course.

Estimation

- Given the training data

Student	ML grade
$\mathbf{x}_1$	A
$\mathbf{x}_2$	B
...	...

  we need to find a mapping from "input vectors" $\mathbf{x}$ to "labels" y encoding the grades for the ML course.

- Possible solution (nearest neighbor classifier):
  1. For any student $\mathbf{x}$ find the "closest" student $\mathbf{x}_i$ in the training set
  2. Predict y_i, the grade of the closest student

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)

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 test sets
  - evaluate the classifier constructed on the basis of only the smaller training set on the new test set

Model selection

- We can refine
  - the estimation algorithm (e.g., using a classifier other than the nearest neighbor classifier)
  - 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

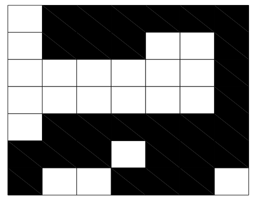
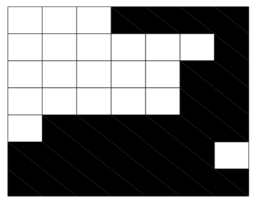
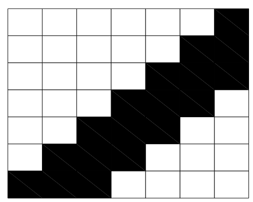
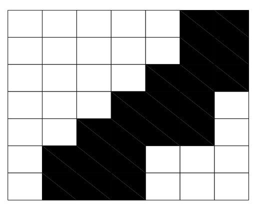
Types of learning problems

A rough (and somewhat outdated) classification of learning problems:

- Supervised learning, where we get a set of training inputs and outputs
  - classification, regression
- Unsupervised learning, where we are interested in capturing inherent organization in the data
  - clustering, density estimation
- Reinforcement learning, where we only get feedback in the form of how well we are doing (not what we should be doing)
  - planning

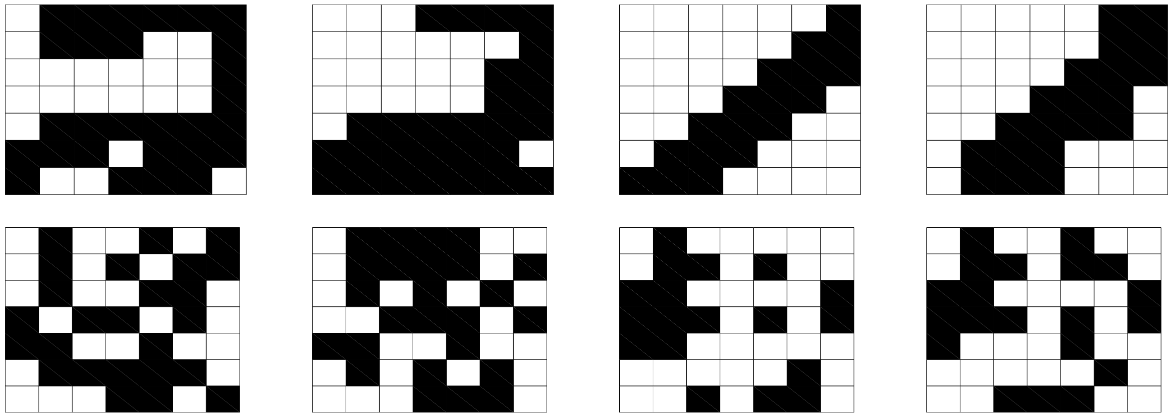
Supervised learning: classification (again)

Example: digit recognition (8x8 binary digits)

binary digit	target label
	“2”
	“2”
	“1”
	“1”
...	...

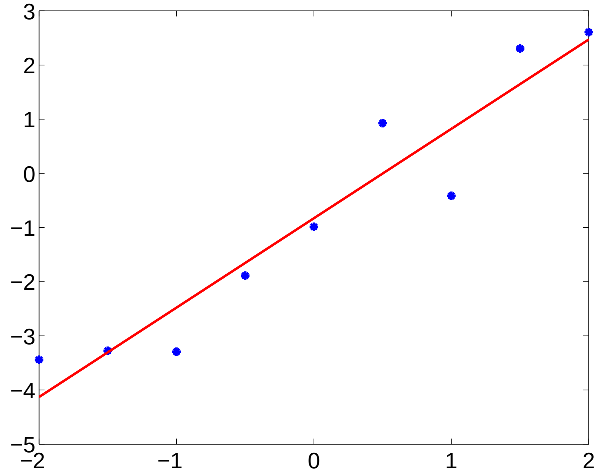
- We wish to learn the mapping from digits to labels

Representation/assumptions



- A change in the representation that preserves the relevant information can preclude learning

Supervised learning: regression

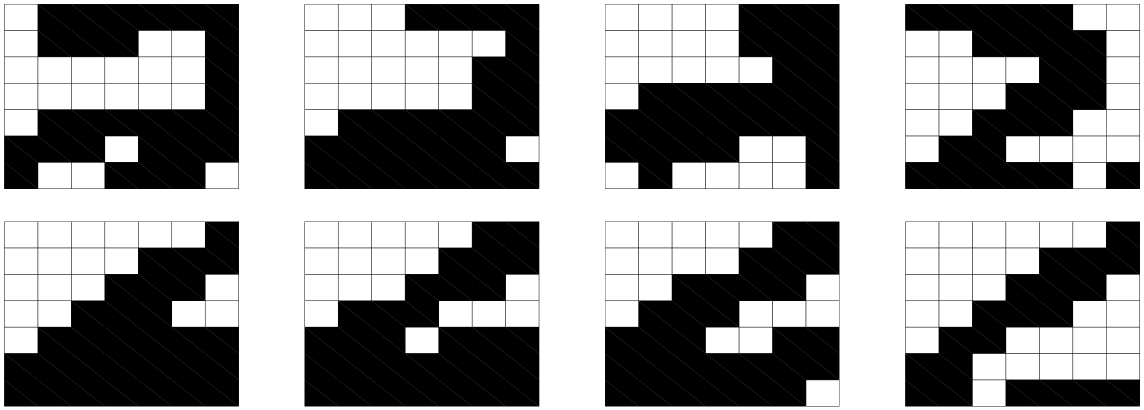


- Given a set of training examples $\{(\mathbf{x}_1, y_1) \ldots, (\mathbf{x}_n, y_n)\}$, we want to learn a mapping $f : \mathcal{X} \rightarrow \mathcal{Y}$ such that

$$y_i \approx f(\mathbf{x}_i), \;\; i = 1, \ldots, n$$

Unsupervised learning: data organization

The digits again...



- We'd like to understand the generation process of examples (digits in this case)