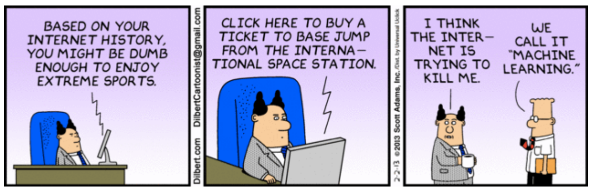
CS545 Machine Learning

Course introduction



1

Machine learning and related fields

Machine learning: the construction and study of systems that learn from data.

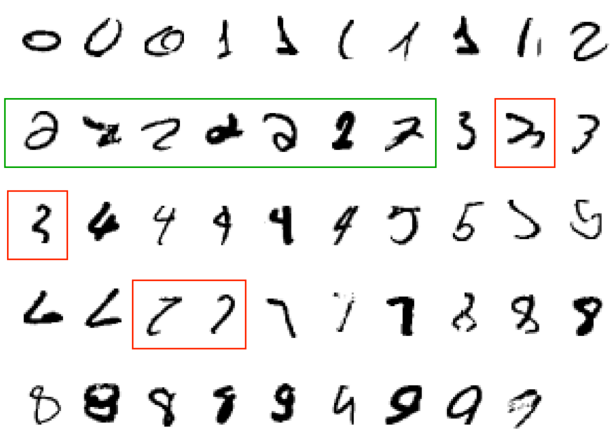
Pattern recognition: the same field, different practitioners

Data mining / big data: using ML to discover patterns in (big) data

Statistics and probability: a lot of algorithms have a probabilistic flavor

2

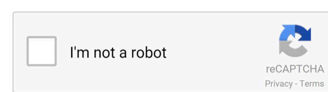
Example problem: handwritten digit recognition



How is this relevant to machine learning?

Remember these?

Now we have:



CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart)

4

Tasks best solved by a learning algorithm

Recognizing patterns and anomalies:
- Face recognition
- Handwritten or spoken words
- Medical images
- Unusual credit card transactions
- Unusual patterns of sensor readings (in nuclear power plants or car engines)
- Stock prices

Examples of machine learning on the web

Spam filtering

Recommendation systems (amazon, netflix):
- Lots of noisy data.

Information retrieval:
- Find documents or images that are relevant to a query.

"The machine learning algorithm wants to know if we'd like a dozen wireless mice to feed the Python book we just bought."

http://www.kdnuggets.com/2014/12/cartoon-unexpected-data-science-recommendations.html

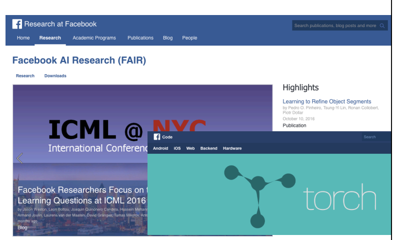
---

TensorFlow

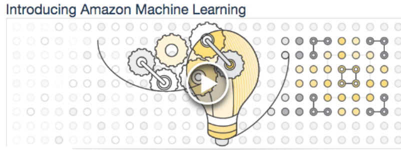
https://www.tensorflow.org/

DeepMind

https://deepmind.com/



Introducing Amazon Machine Learning



https://aws.amazon.com/machine-learning/

News › Business › Business News

Apple buys Turi machine-learning startup for a reported $200m

Turi lets developers and data scientists incorporate machine learning and artificial intelligence into their apps

Zlata Rodionova | Wednesday 10 August 2016 | 2 comments

http://www.independent.co.uk/news/business/news/apple-turi-ai-tim-cook-machine-learning-startup-a7182476.html

---

Google Research Blog

The latest news from Research at Google

Computer, respond to this email.

Tuesday, November 03, 2015

Posted by Greg Corrado*, Senior Research Scientist

Machine Intelligence for You

https://research.googleblog.com/2015/11/computer-respond-to-this-email.html

http://www.kdnuggets.com/2015/11/cartoon-machine-learning-iphone-hr-manager.html

8

Course Objectives

The machine learning toolbox

- Formulating a problem as an ML problem
- Understanding a variety of ML algorithms
- Running and interpreting ML experiments
- Understanding what makes ML work – theory and practice

The Book

Learning from data

http://amlbook.com

The book website contains several e-chapters

Grading

Assignments + project are 95% of the grade

Of that: 5 assignments, worth 80%

- Combination of implementation, running ML experiments, and theory questions

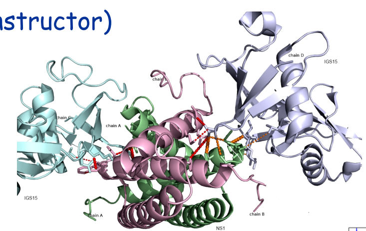
A project assignment worth 20%

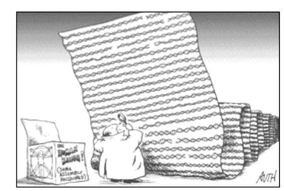
- You choose what you want to work on!

The rest: Canvas quizzes

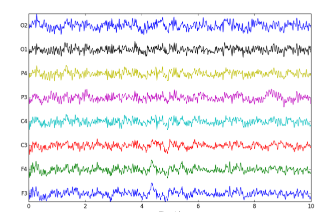
Course staff

Asa Ben-Hur (instructor)





Dejan Markovikj (TA)



About this course

Course webpage:

http://www.cs.colostate.edu/~cs545

Slides/assignments are posted on the course webpage's schedule page.

Canvas will be used for: ~~forums~~, grades, quizzes

Piazza will be our primary communication tool

Implementation: Python

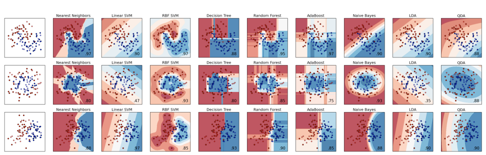
Why Python?

- A concise and intuitive language
- Simple, easy to learn syntax
- Highly readable, compact code
- Supports object oriented and functional programming
- Strong support for integration with other languages (C,C++,Java)

Implementation: Python

Why Python for ML?

- An interpreted language – allows for interactive data analysis
- Libraries for plotting and vector/matrix computation
- Cross-platform compatibility
- Free
- Language of choice for many ML researchers (other options: matlab, R); many ML packages available.



Why Python

I am more productive!

- Machine performance vs. programmer performance

Makes programming fun!

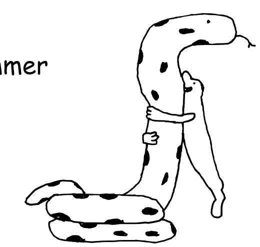


image from: ftp://www.mindview.net/pub/eckel/LovePython.zip

Which version?

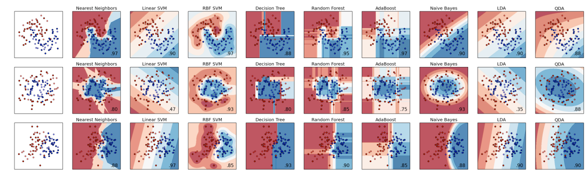
2.x or 3.x?

Python 3 is a non-backward compatible version that removed a few "warts" from the language.

We will use Python 3.

ML in Python

Concepts and algorithms will be demonstrated using scikit-learn
Available at : http://scikit-learn.org/



NumPy: operations on arrays and matrices

Matplotlib: plotting library

The Anaconda python distribution

Contains (almost) all the packages that will be used in this course.

ANACONDA

How will we learn Python?

- Overview of Numpy/scikit-learn in lecture.
- Course website has links to Python tutorials and other resources

Labeled data

E-mail	$feature_1$	$feature_2$	Spam?
1	1	1	1
2	1	0	-1
3	0	1	-1
4	0	0	1
5	0	0	-1

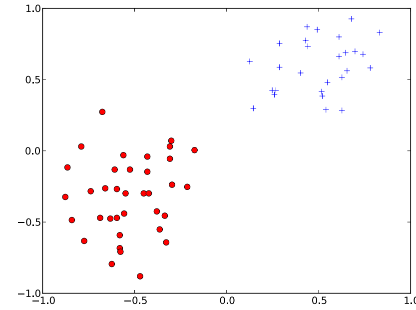
$feature_1$ and $feature_2$ are two characteristics of emails (e.g. the presence of the word "viagara"). These are called features.

Spam? Is the label associated with the each email

This is a binary classification problem

21

Binary classification



Scatter plot of labeled data with two features (dimensions)

22

Another example: Credit approval

Should an applicant be approved for credit?

Feature	Value
Gender	Male
Salary	70,000
Debt	21,000
Years in job	1 year
…	

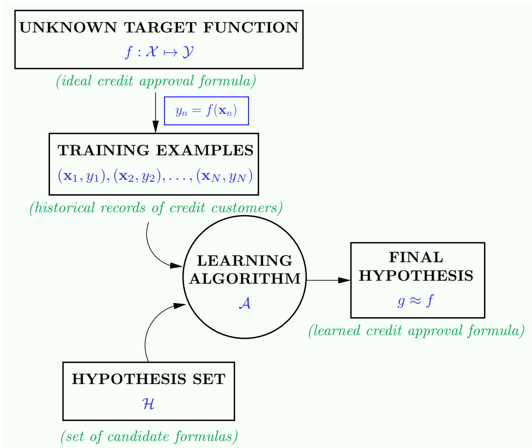
23

Key components of a learning problem

- Salary, debt, years in residence, … — *input* $\mathbf{x} \in \mathbb{R}^d = \mathcal{X}$.
- Approve credit or not — *output* $y \in \{-1, +1\} = \mathcal{Y}$.
- True relationship between $\mathbf{x}$ and y — *target function* $f : \mathcal{X} \mapsto \mathcal{Y}$.
  (The target f is *unknown*.)
- Data on customers — *data set* $\mathcal{D} = (\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)$.
  ($y_n = f(\mathbf{x}_n)$.)

24

The ML setup



25

Not every problem is a ML problem

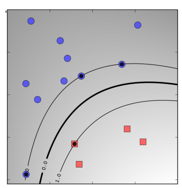
Which of the following problems are best suited for machine learning?

- Classifying numbers into primes and non-primes.
- Predicting the time it would take a falling object to hit the ground.
- Determining the optimal cycle for traffic lights in a busy intersection.
- Medical diagnosis

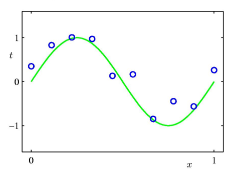
26

ML tasks

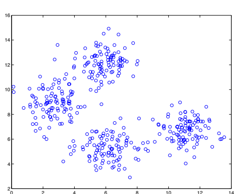
Classification: discrete/categorical labels



Regression: continuous labels



Clustering: no labels

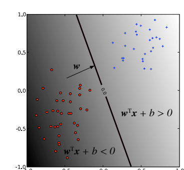


27

Types of models

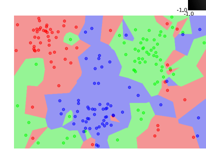
Geometric

- Ridge-regression, SVM, perceptron



Distance-based

- K-nearest-neighbors

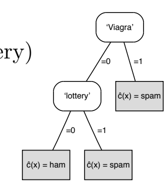


Probabilistic

- Naïve-bayes $P(Y = \text{spam}|\text{Viagara, lottery})$

Logical models: Tree/Rule based

- Decision trees



28

Types of learning tasks

Supervised learning
- Learn to predict output given labeled examples

Unsupervised learning
- Data is unlabeled
- Create an internal representation of the input e.g. form clusters; extract features
- Most "big" datasets do not come with labels

Reinforcement learning
- Maximizing "reward" (not covered).

ML in Practice

- Understanding the domain, and goals
- Creating features, data cleaning and preprocessing
- Learning models
- Interpreting results
- Consolidating and deploying discovered knowledge

An iterative process

Human vs machine learning

Human	Machine
Observe someone, then repeat	Supervised Learning
Keep trying until it works (riding a bike)	Reinforcement Learning
Memorize	k-Nearest Neighbors
20 Questions	Decision Tree
A network of neurons with complex interconnections	Neural networks

Training vs testing

hypothesis target function

Out-of-sample error: (testing error)

$$E_{\text{out}}(h) = \mathbb{P}_{\mathbf{x}}\left[h(\mathbf{x}) \neq f(\mathbf{x})\right]$$

In-sample error: (training error)

$$E_{\text{in}} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\left[h(\mathbf{x}_i) \neq f(\mathbf{x}_i)\right]$$

Indicator function

Training: finding a rule that minimizes E_{in}

Testing: getting an estimate of E_{out}

Training vs testing

The aim of supervised learning is to do well on test data that is not known during training.

We want the learning machine to model the true regularities in the data and to ignore the noise.

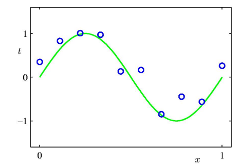
- But the learning machine does not know which regularities are real and which are accidental quirks of the particular set of training examples we happen to have.

So how can we be sure that the machine will generalize well to new data?

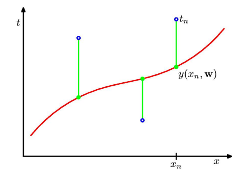
A simple example: Fitting a polynomial

The green curve is the true function (which is not a polynomial)

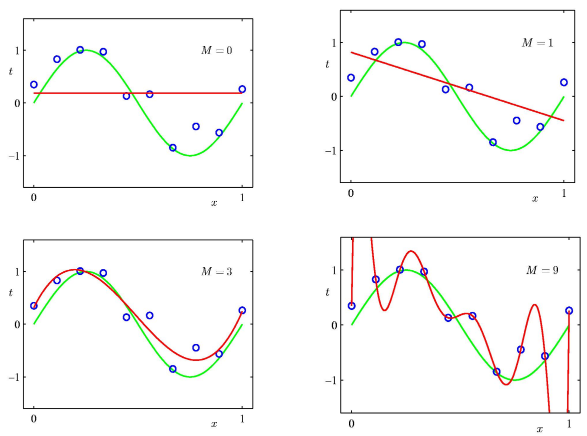
The data points have noise in y.



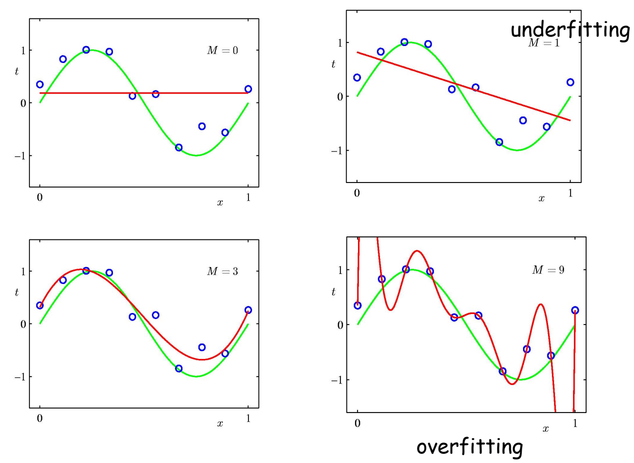
Measure of error (a.k.a. loss function) that measures the deviation of the prediction y(x) from the true value – summed over all the examples.



Which model is best?



Which model is best?

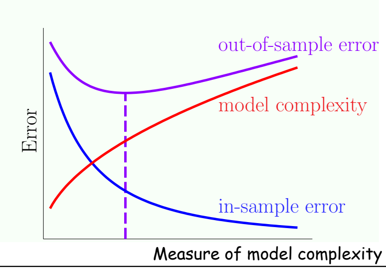


Figures from: "Pattern Recognition and Machine Learning" by Christopher Bishop

Trading off goodness of fit against model complexity

You can only expect a model to generalize well if it explains the data surprisingly well given the complexity of the model.

If the model has as many degrees of freedom as the data, it can fit the data perfectly. But so what?



What we'll cover

Supervised learning

- Linear classifiers
- Support vector machines
- Neural networks and deep learning
- Nearest neighbor classifiers
- Probabilistic classifiers
- Decision trees and ensemble models

Unsupervised learning

- Clustering
- Dimensionality reduction

Running and interpreting ML experiments