Machine Learning - Intro

Barnabas Poczos

Machine Learning 10-715 Sept 9, 2015



Administration

The Team

Instructors:

Barnabas Poczos

- bapoczos@cs.cmu.edu
- office hours after class

Alex Smola

- alex@smola.org
- office hours after class

TAs:

Hsiao-Yu Fish Tung

- sfish0101@gmail.com
- office hours Tue 3:30pm-4:30pm, GHC 8208

Eric Wong

- ericwong@andrew.cmu.edu
- Office hours: TBD

Class Assistant:

Mallory Deptola

- mdeptola@andrew.cmu.edu
- office: GHC 8001

Machine Learning Class webpage

http://www.cs.cmu.edu/~bapoczos/Classes/ ML10715_2015Fall/

Machine Learning

Home

Description

FAQ

Problems

Datasets

Google Group

Lectures

Recitations

CMU

Alex Smola Barnabas Poczos Eric Wong Hsiao Yu Fish Tung MID

Advanced Introduction to Machine Learning

10-715

Practical information

- Lectures: Monday and Wednesday, 10:30AM to 11:50AM, Location: GHC 4102
- Recitations: Tuesdays 5:00PM to 6:00PM, Location: Wean Hall 8427
- Instructor: Barnabas Poczos (office hours after class) and Alex Smola (office hours after class)
- TAs: Hsiao-Yu Fish Tung (office hours Tuesdays 3:30pm-4:30pm in GHC 8208) and Eric Wong (office hours @@@ in GHC @@@)
- Grading Policy: Homework (40%), Midterm (20%), Project (40%).
- Google Group: Join it here. This is the place for announcements.

Updates

Sept 9 2015: Initial site update

Resources

Auditing

To satisfy the auditing requirement, you must

- ☐ Do the homeworks and pass + do the midterm and pass.
- ☐ Please send the instructors and TAs an email saying that you will be auditing the class.

Prerequisites

Probabilities Distributions, densities, marginalization, indepdence... Basic statistics Moments, typical distributions, regression... ☐ Basic algebra: SVD, eigenvectors, orthonormal matrices, ... ☐ Algorithms Dynamic programming, data structures, complexity O() ... Programming Your choice of language, but Matlab will be very useful ☐ We provide some background, but the class will be fast paced ☐ Ability to deal with "abstract mathematical concepts"

Recitations

- Strongly recommended
 - Brush up pre-requisites
 - Review material (difficult topics, clear misunderstandings, extra new topics)
 - Ask questions
- Tuesdays: 5:00PM to 6:00PM, Location: Wean Hall 8427
- 5 special office hours instead of recitations
 - same time and same place as recitations:
 Discussions of homework & midterm solutions

Textbooks

- No required book
- Reading assignments on class homepage
- Recommended Textbook:
 - Pattern Recognition and Machine Learning; Chris Bishop
- Secondary Textbooks:
 - The Elements of Statistical Learning: Data Mining, Inference, and Prediction; Trevor Hastie, Robert Tibshirani, Jerome Friedman
 - Machine Learning; Tom Mitchell
 - Information Theory, Inference, and Learning Algorithms; David MacKay

Grading

4 Homeworks (40%)

- Start early, Sta

Final project (40%)

- Form groups by next week
- Optimal group size is 3
- Proposal, Midterm report, Final report, Group presentations (peer graded)
- Applying machine learning to your research area
- NLP, IR, vision, robotics, computational biology, Outcomes that offer real utility and value

Midterm (20%)

- Mon., Nov 9 in class.

Theory exercises and/or analysis. Dates already set (no "ticket already booked", "I am in a conference", etc. excuse ...)

Homeworks

- ☐ Homeworks are hard, start early ⓒ
- ☐ Due in the beginning of class
- 2 late days for the semester
- ☐ After late days are used up: zero credit
- ☐ Submissions: hard copy in the beginning of class + email to TAs

Homeworks

Collaboration

- You may discuss the questions
- Each student writes their own answers
 - ... copying from whiteboard is not acceptable!
- Each student must write their own code for the programming part
 - ... simply renaming variables is not acceptable!
- Please don't search for answers on the web, Google, previous years' homeworks, etc.
 - please ask us if you are not sure if you can use a particular reference

First Point of Contact for HWs

 To facilitate interaction, a TA will be assigned to each homework question

 This will be your "first point of contact" for this question

Communication Channel

For announcements, subscribe to the Google group:

https://groups.google.com/d/forum/10-715-fall-2015-cmu

Meetings with Barnabas

☐ Office hours

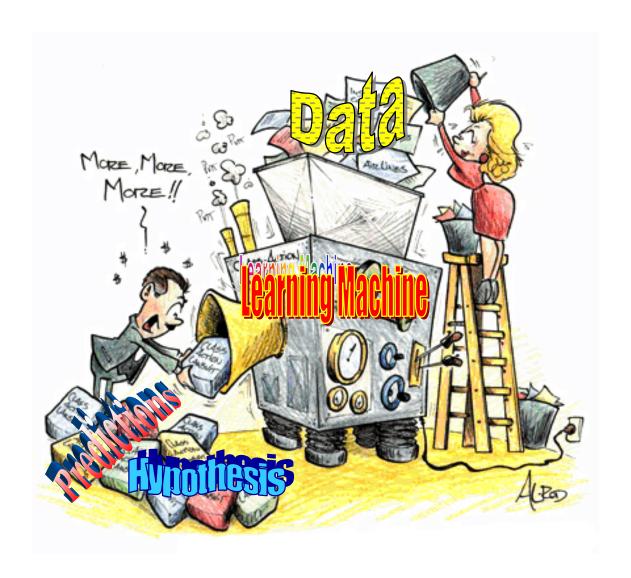
Or

☐ Email Barnabas's assistant, Sandy Winkler: sandyw@cs.cmu.edu to schedule a meeting.

Any other questions about administration and logistics?

What is Machine Learning?

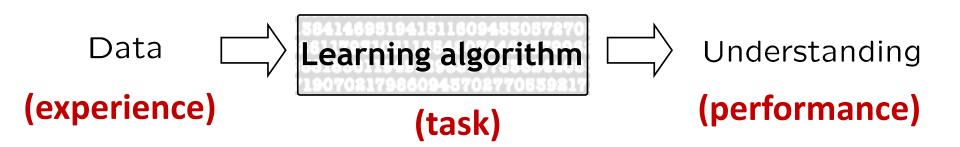
What is Machine Learning?



What is Machine Learning?

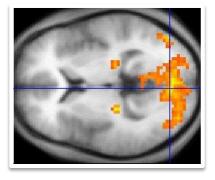
Study of algorithms that

- improve their <u>performance</u>
- at some task
- with <u>experience</u>



From Data to Understanding ... Machine Learning in Action

Decoding thoughts from brain scans





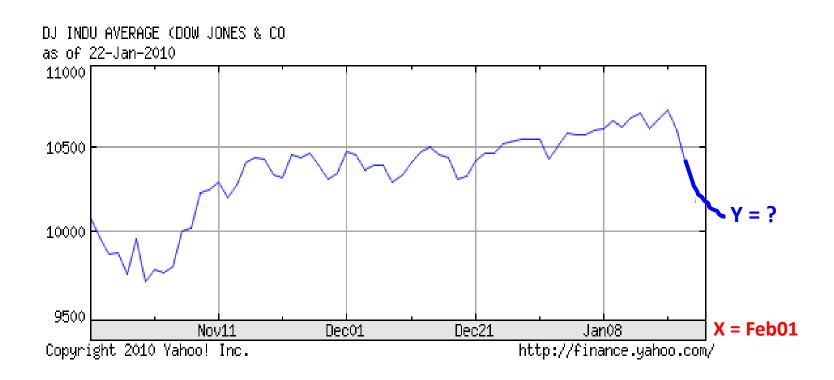
Rob a bank ...





The average Joe's MRI scan can show a brain abnormality, do we proceed to check him into the nearest mental institution or prison? That would make about as much sense as trying to prove a defendant innocent of a violent

Stock Market Prediction



Document classification



Spam filtering

Welcome to New Media Installation: Art that Learns

Hi everyone,

Welcome to New Media Installation: Art that Learns

The class will start tomorrow.

Make sure you attend the first class, even if you are on the Wait List. The classes are held in Doherty Hall C316, and will be Tue, Thu 01:30-4:20 PM.

By now, you should be subscribed to our course mailing list: 10615-announce@cs.cmu.edu.

Natural LoseWeight SuperFood Endorsed by Oprah Winfrey, Free Trial 1 bottle, pay only \$5.95 for shipping mfw rlk | Spam | X



Spam/ Not spam



=== Natural WeightL0SS Solution ===

Vital Acai is a natural WeightLOSS product that Enables people to lose wieght and cleansing their bodies faster than most other products on the market.

Here are some of the benefits of Vital Acai that You might not be aware of. These benefits have helped people who have been using Vital Acai daily to Achieve goals and reach new heights in there dieting that they never thought they could.

- * Rapid WeightL0SS
- * Increased metabolism BurnFat & calories easily!
- * Retter Mood and Attitude

Cars navigating on their own



Boss, the self-driving SUV
1st place in the DARPA Urban
Challenge.

Photo courtesy of Tartan Racing.



Many, many more...

```
Speech recognition, Natural language processing
Computer vision
Medical outcomes analysis
Computational biology
Sensor networks
Social networks
Robocup
...
```

25

ML is trending!

- Wide applicability
- Study very large-scale complex systems
 - Internet (billions of nodes), sensor network (new multi-modal sensing devices), genetics (human genome)
- Huge multi-dimensional data sets
 - 30,000 genes x 10,000 drugs x 100 species x ...
- Improved machine learning algorithms
- Improved data capture (Terabytes, Petabytes of data),
- faster computers , faster network

Machine Learning Tasks

Broad categories -

Supervised learning

Classification, Regression

Unsupervised learning

Density estimation, Clustering, Dimensionality reduction

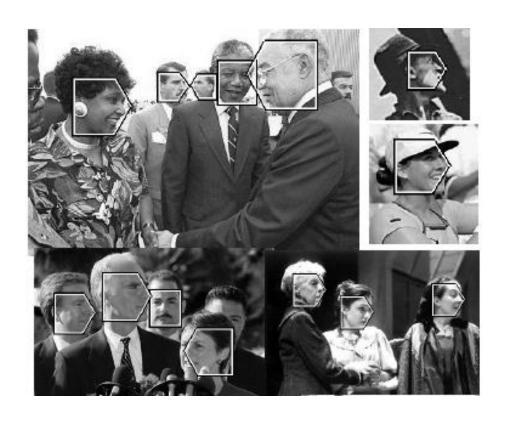
- Semi-supervised learning
- Active learning
- Reinforcement learning
- Online learning
- Transfer learning
- Multitask learning
- Many more ...

Supervised Learning problems

Features?

Labels?

Classification/Regression?



Face Detection

Supervised Learning problems

Features?

Labels?

Classification/Regression?

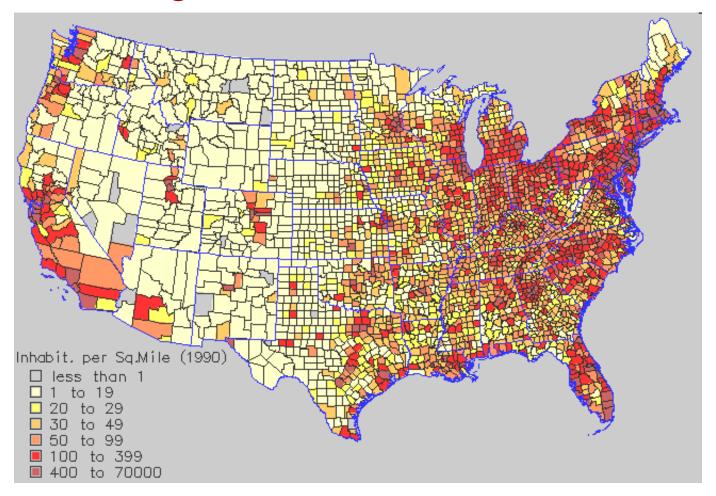


Robotic Control

Unsupervised Learning – Density Estimation

Population density

Aka "learning without a teacher"



Unsupervised Learning – clustering

Group similar things e.g. images

[Goldberger et al.]





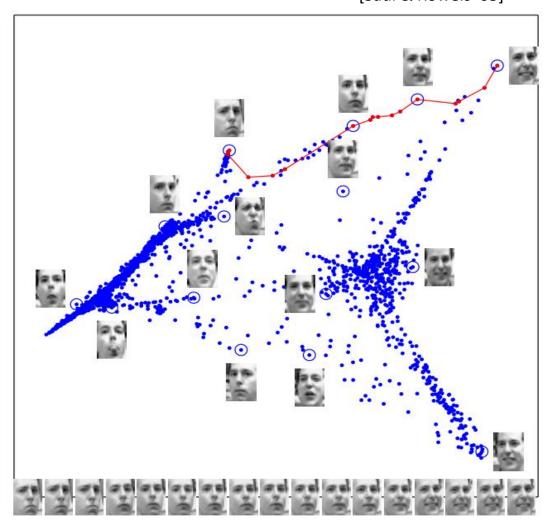
Unsupervised Learning - Embedding

Dimensionality Reduction

[Saul & Roweis '03]

Images have thousands or millions of pixels.

Can we give each image a coordinate, such that similar images are near each other?



Unsupervised Learning - Embedding

```
billmark
                                         mary
              bob jack
                                           elizabeth
                                 stephen
                                       edward
              jimmikerianichpellercharbenrelexander
evenris
                         andrew josáním josánímcismaria
frankaviániejames louis
            joe tom harry
       mr.
                          arthur<sub>george</sub> j
thomas
             don
                    ray
                                martin
                                       howard
                            simon
                 ben
                                          lee
                \mathbf{al}
                                    scott
   dr.
                                       lewis bush
                                       tayIorjokiiskiionEox
                                          davis
                                                    ford grant
                                                    bell.
  von
      van
                         discongeles
                 los
                   santa
 des
              hong
core
                                                                   juneaugust
```

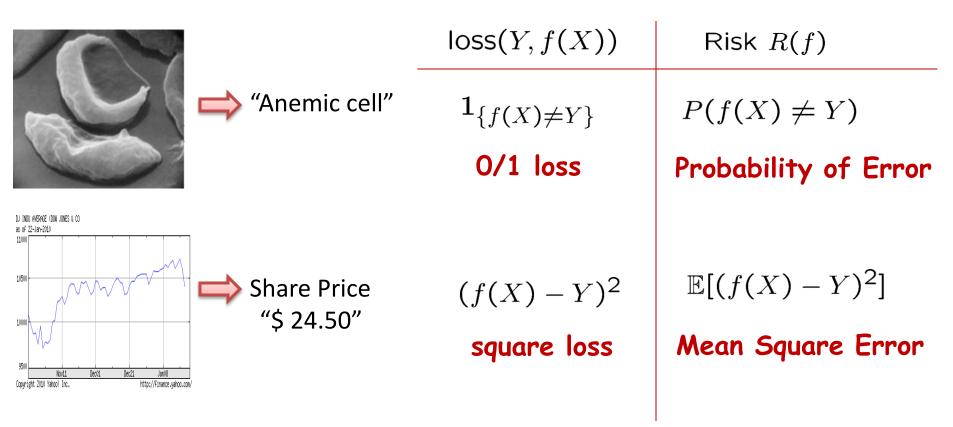
cape

SALCERT

```
columbia
indiamturissouri
marviand
                colorado termessee
washitgkas oregilli lans Garo
california in 1888 caro
      holly bid ago toronto ontarias sachusetts your toronto
                                              caribrildge
                       manchester
               london
                                    victoria
            bep<u>hins</u>
                            quebec
               MOSCOW
                                     scotland
                          mexico
                                  walangland
                                 ireland britain
                    canada
                       aus tankii gweden
             singapore america norwatiance
                   asia russia austr
asiarica russia
                           europe
ankong
                       indiajapan rome
                       pak China
usa philippinds
```

Performance Measures

Performance: Risk $R(f) \equiv \mathbb{E}_{XY} [loss(Y, f(X))]$



Bayes Optimal Rule

<u>Ideal goal</u>: Construct prediction rule $f^*: \mathcal{X} \to \mathcal{Y}$

$$f^* = \arg\min_{f} \mathbb{E}_{XY} [loss(Y, f(X))]$$

Bayes optimal rule

Best possible performance:

Bayes Risk
$$R(f^*) \leq R(f)$$
 for all f

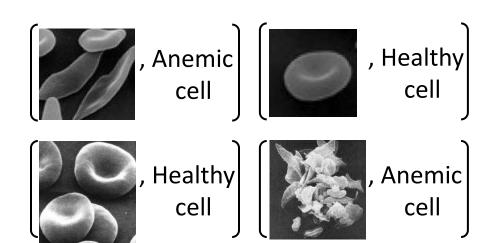
BUT... Optimal rule is not computable - depends on unknown Pxy!

Experience - Training Data

Can't minimize risk since P_{XY} unknown!

Training data (experience) provides a glimpse of P_{XY}

(observed)
$$\{(X_i,Y_i)\}_{i=1}^n \overset{i.i.d.}{\sim} P_{XY}$$
 (unknown) \rightarrow independent, identically distributed



Provided by expert, measuring device, some experiment, ...

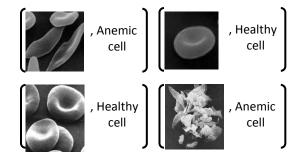
Supervised Learning

Task: Given $X \in \mathcal{X}$, predict $Y \in \mathcal{Y}$.

 \equiv Construct **prediction rule** $f: \mathcal{X} \rightarrow \mathcal{Y}$

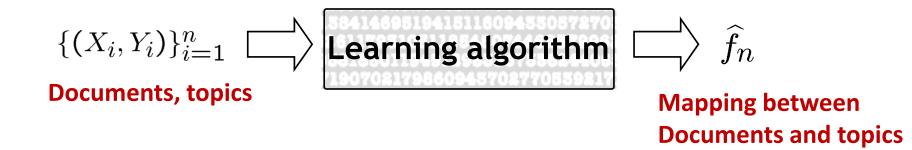
Performance: Risk $R(f) \equiv \mathbb{E}_{XY} \left[loss(Y, f(X)) \right]$ $(X, Y) \sim P_{XY}$

Experience: Training data $\{(X_i, Y_i)\}_{i=1}^n \stackrel{i.i.d.}{\sim} P_{XY}$ (unknown)



Supervised vs. Unsupervised Learning

Supervised Learning – Learning with a teacher



Unsupervised Learning – Learning without a teacher

