Lecture 1: What is Machine Learning?

STAT161/261 Introduction to Pattern Recognition and Machine Learning

Spring 2018

Prof. Allie Fletcher

Lecture 1 Outline

- Course information and details
- What and why machine learning?
- Supervised Learning
  - Examples
  - Classification
  - Regression
- Unsupervised Learning
- Reinforcement Learning
- Why now?

Course Info (see web; most significant bits here)

- We will be using CCLE, after enrollment settles down
- Instructor: Allie Fletcher
- Required Books: *Introduction to Machine Learning* by Ethem Alpaydin and *Pattern Recognition and Machine Learning* by Christopher Bishop
- The majority of what is important will be covered in lectured. However, you will be required to know readings, website handouts, and lecture--not just lecture
- Lecture notes may be slides and handwritten--union of both important :)

What is Machine Learning?

- Learn to improve algorithms from data
- Optimize a performance criterion using example data or past experience

- Role of Statistics: Inference from a sample
- Role of Computer science: Efficient algorithms to
  - Solve the optimization problem
  - Representing and evaluating the model for inference

Why "Learn" ?

- Machine learning is programming computers to optimize a performance criterion using example data or past experience
- There is no need to "learn" to calculate payroll
- Learning is used when:
  - Human expertise does not exist (navigating on Mars)
  - Humans are unable to explain their expertise (speech recognition)
  - Solution changes in time (routing on a computer network)
  - Solution needs to be adapted to particular cases (user biometrics)

What We Talk About When We Talk About "Learning"

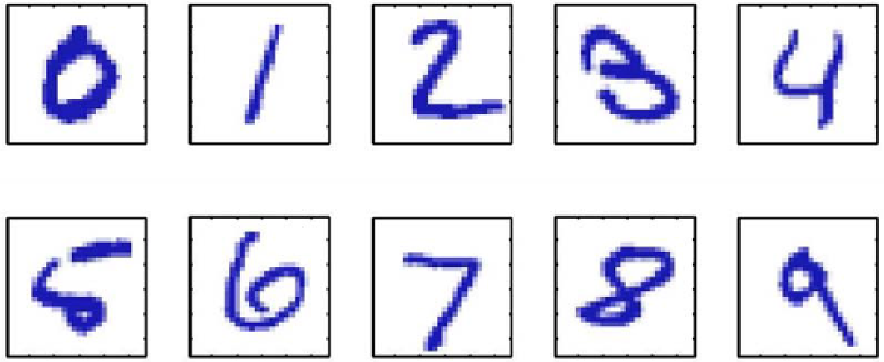
- Learning general models from a data of particular examples
- Data is cheap and abundant (data warehouses, data marts); knowledge is expensive and scarce
- Example in retail: Customer transactions to consumer behavior:

  People who bought "Blink" also bought "Outliers" (www.amazon.com)
- Build a model that is *a good and useful approximation* to the data

Lecture 1 Outline

- Course information and details
- What is machine learning?
- Why do machine learning?
- Supervised Learning
  - Examples
  - Classification
  - Regression
- Unsupervised Learning
- Reinforcement Learning
- Why now?

Example 1: Digit Recognition



Images are 28 x 28 pixels

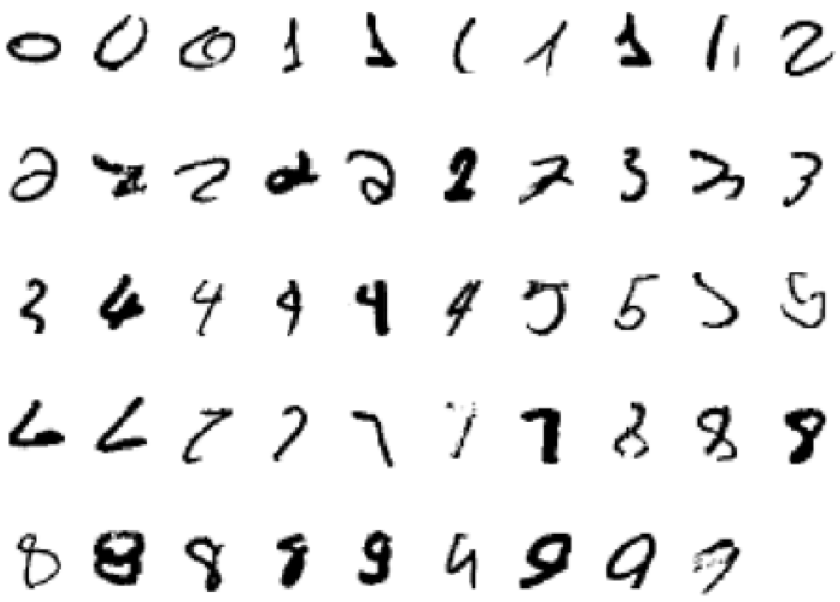
- Recognize a digit from the image
- Learn a function $f(\boldsymbol{x}) \in \{0,1,\ldots,9\}$, $\boldsymbol{x}$ is a 28 x 28 matrix
- Expert systems do not work well:
  - You can recognize the digits, but difficult to program a function $f(\boldsymbol{x})$ that works well
  - Try it!

Supervised Learning on Handwritten Digits

- Supervised:
  - Start with training data, labelled data
- Ex: 6000 examples of each digit
- Learn for example classifier $f(\boldsymbol{x})$ that matches label well on training data
- Given new data $\boldsymbol{x}$ use function to guess digit
- Current systems get <0.21% errors (as of 1/20/2018)

http://rodrigob.github.io/are_we_there_yet/build/classification_datasets_results.html#4d4e495354

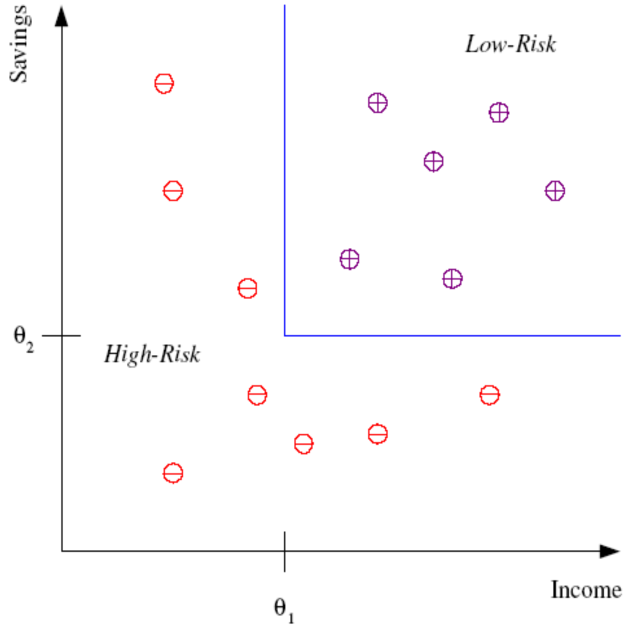
- First commercial application:
  - Used by USPS for recognizing zip codes on letters



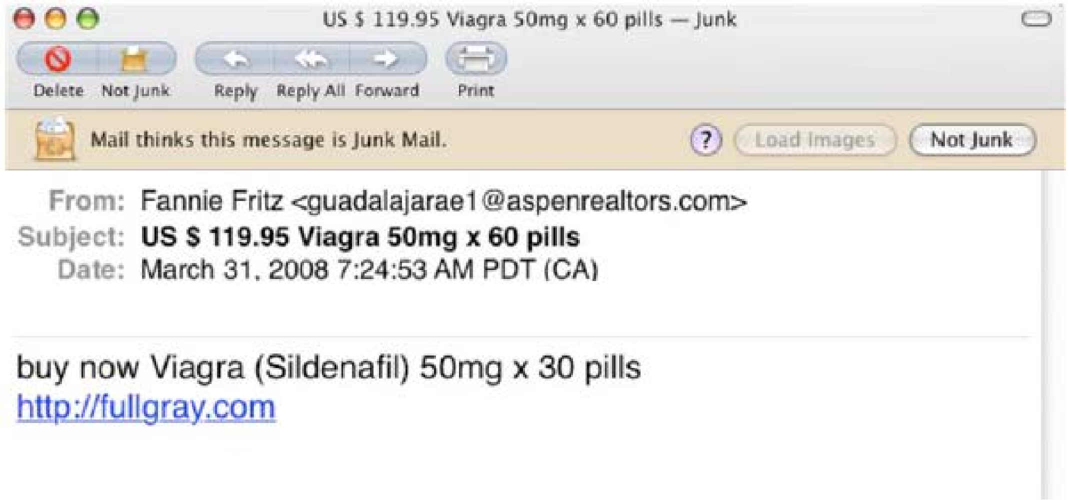
Training examples
Each sample must be labeled by hand who knows truth

Example 2: Credit Score and Classification

- Example: Credit score
- Determine/classify if customer is high-risk or low-risk
- Select some features:
  - Example: income & savings
  - Represent as a vector $x = (x_1, x_2)$
- Learn a function from features to target
  - Use past training data
  - Need to get this data
- The function on the right is an example of a decision tree. If savings are above a line, and then if income is above a line, then the candidate is low-risk.

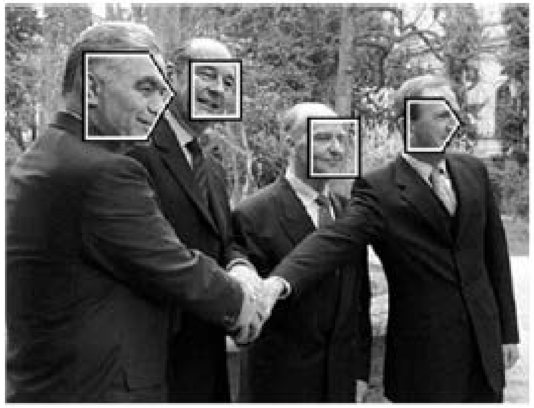
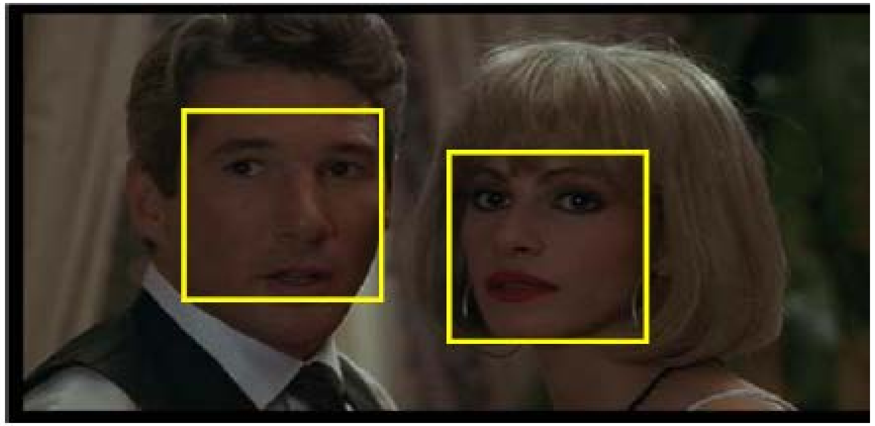


Example 3: Spam Detection



- Classification problem:
  - Is email junk or not junk?
- For ML, must represent email numerically
  - Common model: bag of words
  - Enumerate all words, $i = 1, \ldots, N$
  - Represent email via word count x_i = num instances of word i
- Challenge:
  - Very high-dimensional vector
  - System must continue to adapt (keep up with spammers)

Example 4: Face Detection

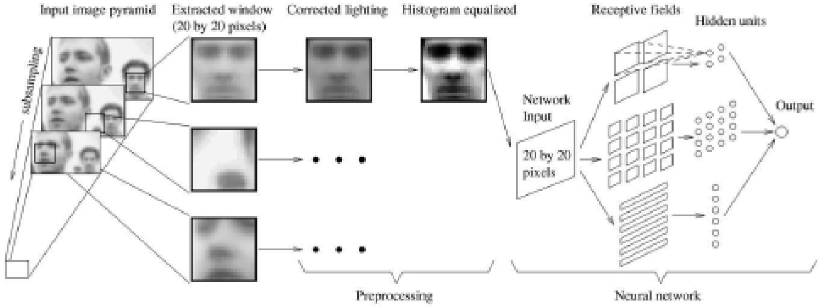
 

- Also a supervised learning problem
- For each image region, determine if
  - Face or non-face

Training Data

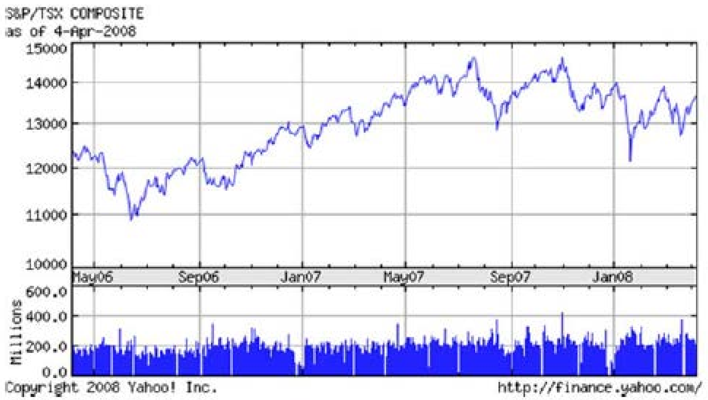
- Typical early face recognition datasets:
- 5000 faces
  - All near frontal
  - Vary age, race, gender, lighting
- 10^8 non faces
- Faces are normalized (scale, translation)
- "functions" that work well may be very complex

- Many more datasets are available now:
  - See http://www.face-rec.org/databases/
  - You can use this for your project!



Rowley, Baluja and Kanade, 1998

Example 5: Stock Price Prediction



- Can you predict the price of a stock?
- What variables would you use?
- What is a non-machine learning approach?

Supervised Learning in General

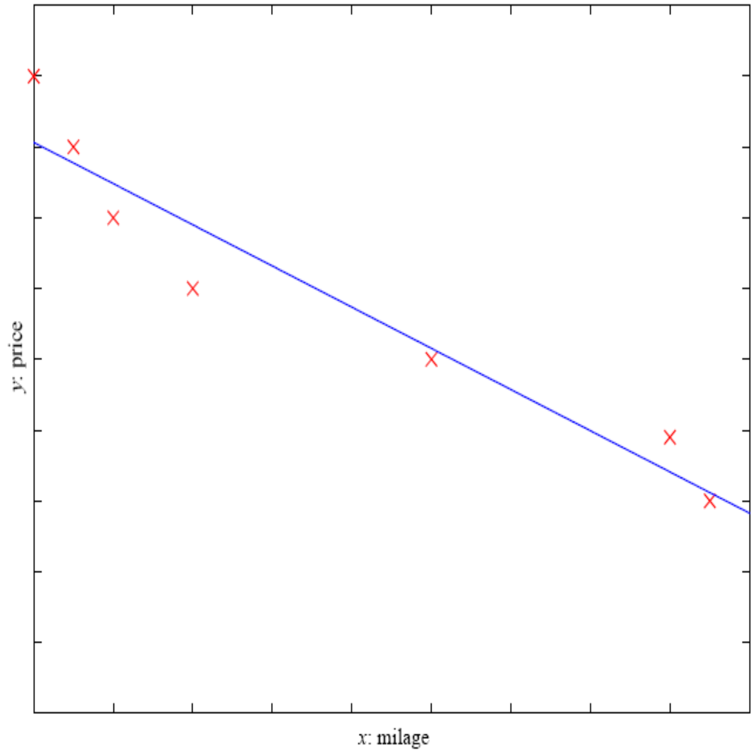
- Prediction of future cases: Use the rule to predict the output for future inputs
- Knowledge extraction: The rule is easy to understand
- Compression: The rule is simpler than the data it explains
- Outlier detection: Exceptions that are not covered by the rule, e.g., fraud

Classification and SL: Many Applications

- Aka Pattern recognition
- Face recognition: Pose, lighting, occlusion (glasses, beard), make-up, hair style
- Character recognition: Different handwriting styles.
- Speech recognition: Temporal dependency.
- Medical diagnosis: From symptoms to illnesses
- Biometrics: Recognition/authentication using physical and/or behavioral characteristics: Face, iris, signature, etc
- …

Regression

- Target variable y is continuous-valued
- Example:
  - Predict $y =$ price of car
  - From $x =$ mileage, size, horsepower, ..
  - Can use multiple predictors
- Assume some form of the mapping
  - Ex. Linear: $y = \beta_0 + \beta_1 x$
  - Find parameters β_0, β_1 from data
- Note: predictors need not be cnts



Regression Example

Machine Learning Repository

Center for Machine Learning and Intelligent Systems

Diabetes Data Set

Download: Data Folder, Data Set Description

File Names and format:
(1) Date in MM-DD-YYYY format
(2) Time in XX:YY format
(3) Code
(4) Value

The Code field is deciphered as follows:

33 = Regular insulin dose
34 = NPH insulin dose
35 = UltraLente insulin dose
48 = Unspecified blood glucose measurement
57 = Unspecified blood glucose measurement
58 = Pre-breakfast blood glucose measurement
59 = Post-breakfast blood glucose measurement
60 = Pre-lunch blood glucose measurement
61 = Post-lunch blood glucose measurement
62 = Pre-supper blood glucose measurement
63 = Post-supper blood glucose measurement
64 = Pre-snack blood glucose measurement
65 = Hypoglycemic symptoms
66 = Typical meal ingestion
67 = More-than-usual meal ingestion
68 = Less-than-usual meal ingestion
69 = Typical exercise activity
70 = More-than-usual exercise activity
71 = Less-than-usual exercise activity

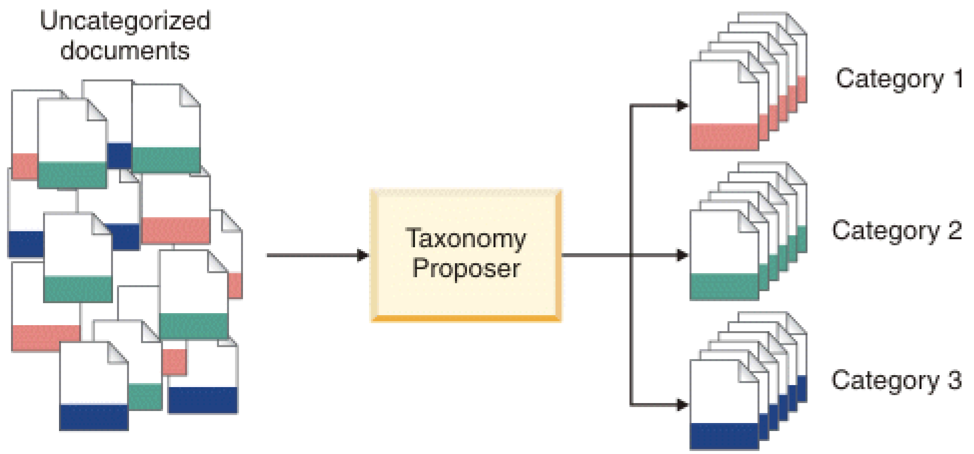
- Predict blood glucose level
- Many possible predictors:
  - Recent past levels
  - Insulin dose
  - Time of last meal
  - …
- Check out data in: https://archive.ics.uci.edu/ml/datasets/Diabetes

Lecture 1 Outline

- Course information and details
- What is machine learning?
- Why do machine learning?
- Supervised Learning
  - Examples
  - Classification
  - Regression
- Unsupervised Learning
- Reinforcement Learning
- Why now?

Unsupervised Learning

- Learning "what normally happens"
- No output
- Clustering: Grouping similar instances
- Example applications
  - Customer segmentation
  - Image compression: Color quantization
  - Bioinformatics: Learning motifs



Example: Document classification
http://www.ibm.com/support/knowledgecenter/SSBRAM_8.7.0/com.ibm.classify.ccenter.doc/c_WBG_Taxonomy_Proposer.htm

Reinforcement Learning

- Learning a policy: A sequence of outputs
- No supervised output but delayed reward
- Credit assignment problem
- Game playing
- Robot in a maze
- Multiple agents, partial observability, …

What ML is Doing Today?

- Autonomous driving
- Jeopardy
- Very difficult games: Alpha
- Machine translation

- Many, many others…

Why Now?

- Machine learning is an old field
  - Much of the pioneering statistical work dates to the 19
- So what is new now?
- Big Data:
  - Massive storage. Large data centers
  - Massive connectivity
  - Sources of data from Internet and elsewhere
- Computational advances
  - Distributed machines, clusters
  - GPUs and hardware

Google Tensor Processing Unit (TPU)

Resources: Journals

- Journal of Machine Learning Research www.jmlr.org
- Machine Learning
- Neural Computation
- Neural Networks
- IEEE Trans on Neural Networks and Learning Systems
- IEEE Trans on Pattern Analysis and Machine Intelligence
- Journals on Statistics/Data Mining/Signal Processing/Natural Language Processing/Bioinformatics/…

Resources: Conferences

- International Conference on Machine Learning (ICML)
- European Conference on Machine Learning (ECML)
- Neural Information Processing Systems (NIPS)
- Uncertainty in Artificial Intelligence (UAI)
- Computational Learning Theory (COLT)
- International Conference on Artificial Neural Networks (ICANN)
- International Conference on AI & Statistics (AISTATS)
- Knowledge Discovery and Data Mining (KDD)
- International Conference on Computer Vision and Pattern Recognition (CVPR)
- International Conference on Computer Vision (ICCV)
- European Conference on Computer Vision (ECCV)

Machine Learning in Almost All Fields

- Retail: Market basket analysis, Customer relationship management (CRM)
- Finance: Credit scoring, fraud detection
- Manufacturing: Control, robotics, troubleshooting
- Medicine: Medical diagnosis
- Telecommunications: Spam filters, intrusion detection
- Bioinformatics: Motifs, alignment
- Web mining: Search engines
- …

Objectives

- Provide examples of machine learning used today
- Given a new problem, qualitatively describe how machine learning can be used
  - Formulate a potential machine learning task
  - Identify the data needed for the task
  - Identify objectives
- Classify a machine learning task:
  - Supervised vs. unsupervised, regression vs. classification
- For supervised learning, identify the predictors and target variables
- Determine the role of expert knowledge in the task vs. data-driven learning