# Lecture 1
# What is Machine Learning?

EE4563/ EL9123: INTRODUCTION TO MACHINE LEARNING

PROF. SUNDEEP RANGAN

# Learning 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

# Outline

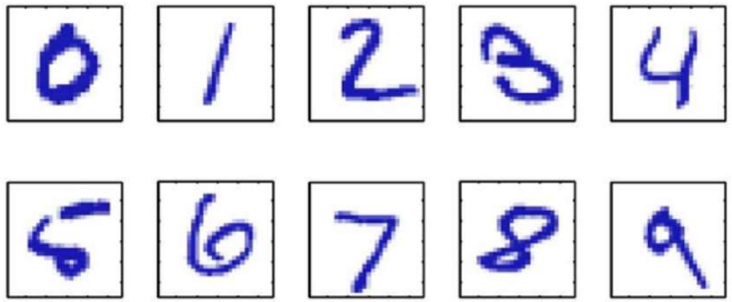
- What is Machine Learning?
- Types of machine learning algorithms
  - Classification
  - Regression
  - Unsupervised learning
- Why the hype today?
- Some slides from:
  - A. Zisserman, “Machine Learning Introduction”
  - Alpaydin, “Introduction to Machine Learning”

# What is Machine Learning?

- Learn to improve algorithms from data.

- Why?
  - 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)

# Example 1: Digit Recognition

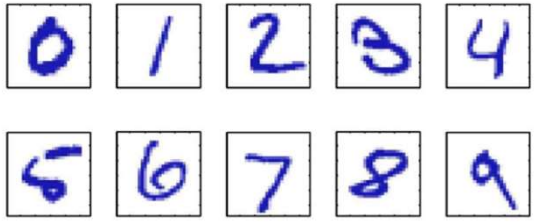


Images are 28 x 28 pixels

- Recognize a digit from the image
- Want a function $f(\boldsymbol{x}) \in \{0,1,\ldots,9\}$, $\boldsymbol{x}$ is a 28 x 28 matrix
  - Takes input as an image
  - Returns the estimated digit

# Classical "Expert" Approach

- Idea: Use your knowledge about digits
  - You are an "expert" since you can do the task
  - Construct simple rules and code them
- Expert rule example: *"Image is a digit 7 if…"*:
  - There is a single horizontal line, and
  - There is a single vertical line
- But, very difficult to make work in practice
  - Lacks robustness
  - Rotations, curves in lines, poorly drawn digits, …
- Problem: We cannot easily code our knowledge
  - Hard to translate to simple mathematical formula

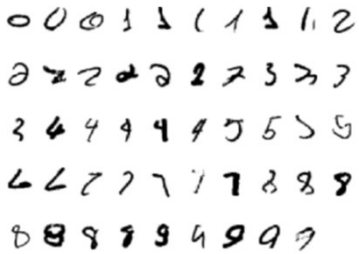


Images are 28 x 28 pixels

```
def count_vert_lines(image):
    ...
def count_horiz_lines(image):
    ...

def classify(image):
    ...
    nv = count_vert_lines(image)
    nh = count_horiz_lines(image)
    ...

    if (nv == 1) and (nh == 1):
        digit = 7
    ...

    return digit
```

# ML Approach: Learn from Data

Training inputs images $\boldsymbol{x}_i$ (ex. 5000 ex per class)



?

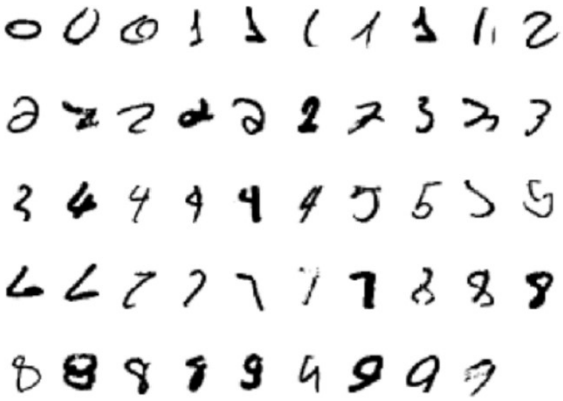
Learned classifier
$f(\boldsymbol{x})$

Training output labels $y_i \in \{0,1,\ldots,9\}$

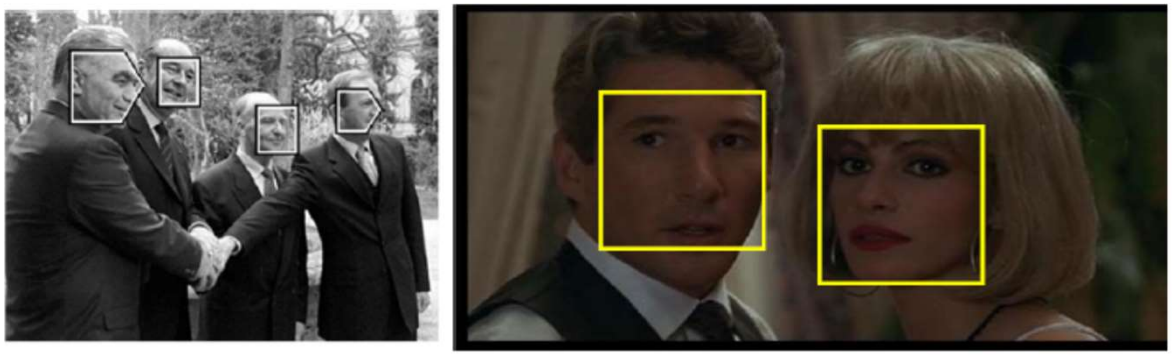
- Supervised learning:
  - Get many labeled examples $(\boldsymbol{x}_i, y_i), i = 1,\ldots,N$ (Called the training data)
  - Each example has an input $\boldsymbol{x}_i$ and output $y_i$
  - Learn a function $f(\boldsymbol{x})$ such that: $f(\boldsymbol{x}_i) = y_i$ for "most" training examples
  - Use this function on new $\boldsymbol{x}$
- No manual coding!

# ML Approach Challenges

- Learned systems do very well on image recognition problems
  - On digit recognition, current systems get <0.21% errors (as of 1/20/2018)
  - Used widely in commercial systems today (e.g. OCR)
  - Cannot match this performance with an expert system
- But, there are challenges:
  - Need labeled data. Someone has to manually create this.
  - How do we search over a set of functions $f(x)$?
  - If a function works on training example, will it work on new data?
- Some things you will learn in this course
  - How to parametrize a set functions
  - How to fit a function
  - How to ensure it generalizes to new examples



# Example 2: Face Detection

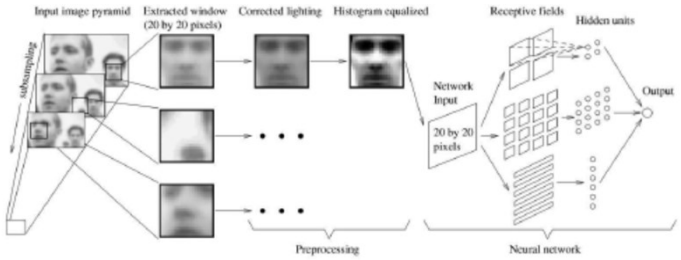


- Also a supervised learning problem
- For each image region, determine if
  - Face or non-face

# Training Data

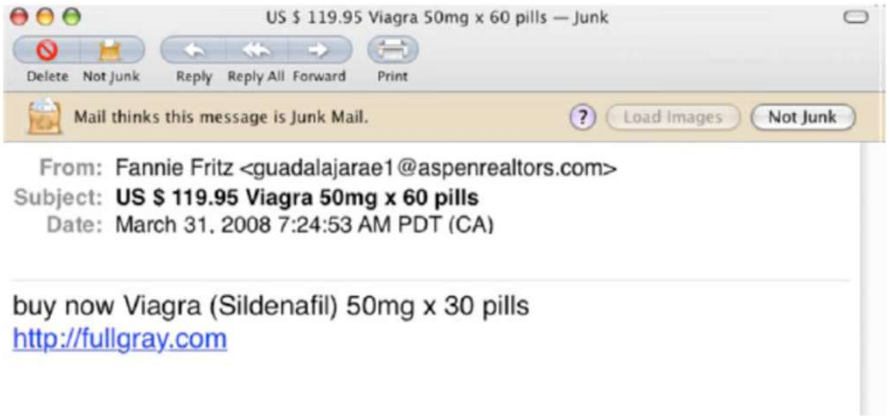
- Typical early face recognition datasets:
- 5000 faces
  - All near frontal
  - Vary age, race, gender, lighting
- 10^8 non faces
- Faces are normalized (scale, translation)
- "functions" that work well may be very complex

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



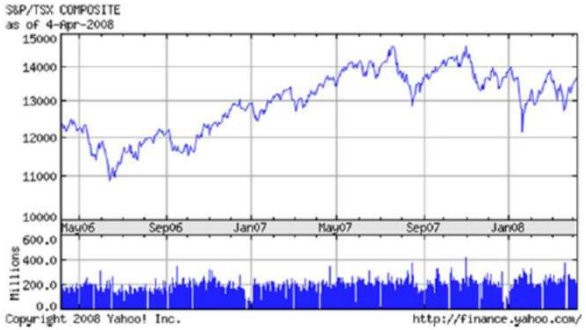

Rowley, Baluja and Kanade, 1998

# Example 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: Stock Price Prediction




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

# Machine Learning in Many 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
- ...

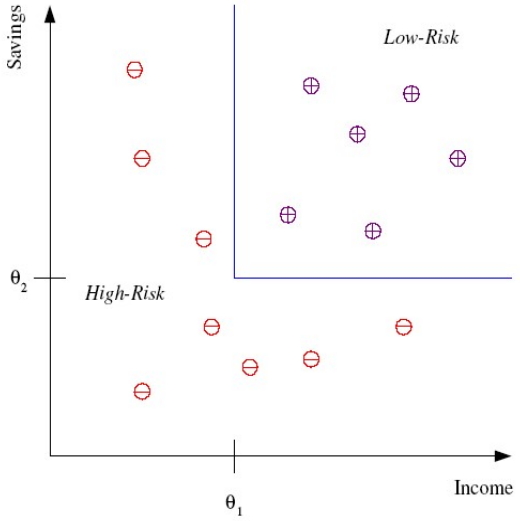
# Outline

- What is Machine Learning?
- Types of machine learning algorithms
  - Classification
  - Regression
  - Unsupervised learning
- Why the hype today?

- Some slides from:
  - A. Zisserman, “Machine Learning Introduction”
  - Alpaydin, “Introduction to Machine Learning”

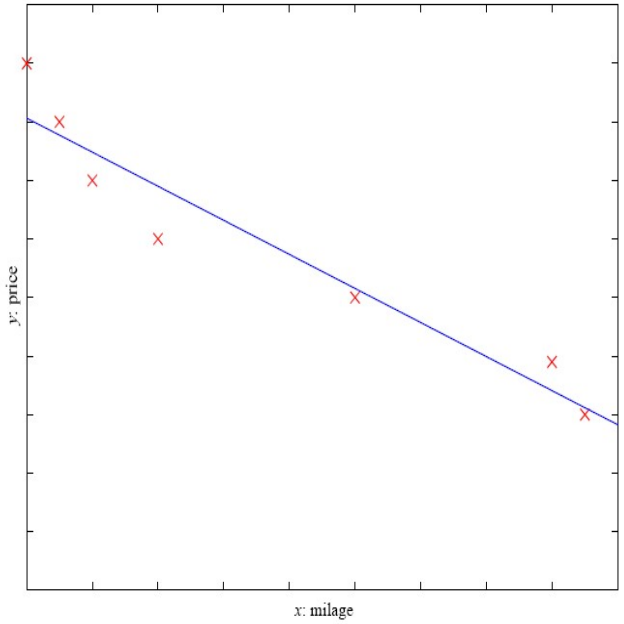
# Classification

- Example: Credit score
- Determine 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.

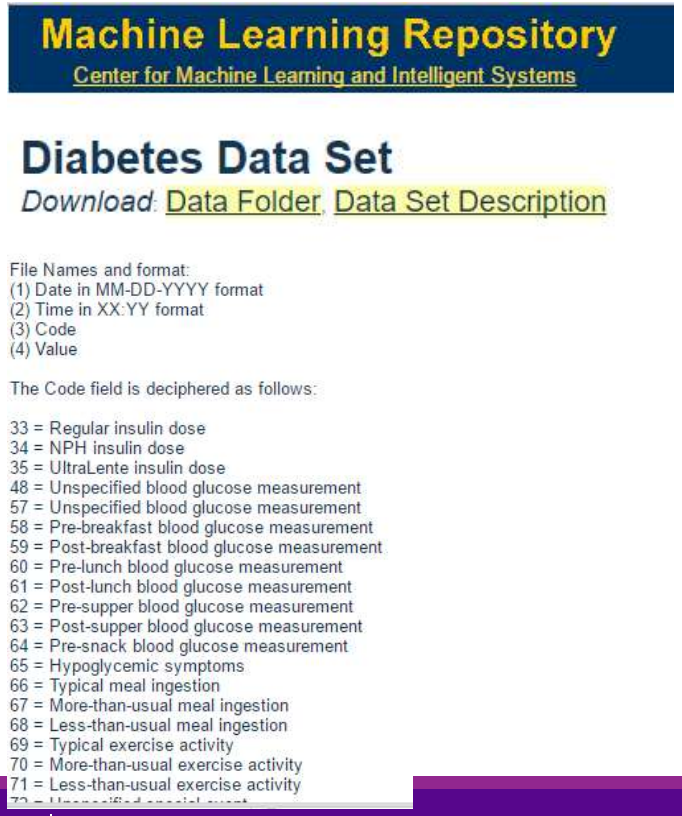


# Regression

- 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 $\beta_0, \beta_1$ from data



# 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

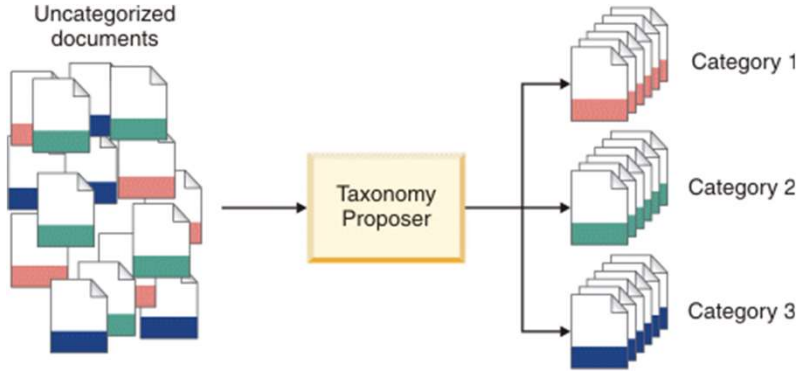
The Code field is deciphered as follows:

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

# 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

# Outline

- What is Machine Learning?
- Types of machine learning algorithms
  - Classification
  - Regression
  - Unsupervised learning
- Why the hype today?
- Some slides from:
  - A. Zisserman, “Machine Learning Introduction”
  - Alpaydin, “Introduction to Machine Learning”

# What ML is Doing Today?

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

- Many, many others…



# Why Now?

- Machine learning is an old field
  - Much of the pioneering statistical work dates to the 1950s
- 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)

# Top 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/...

# Top 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)

# Exercise

- Break into small groups
- Take a field that interests you:
  - Ex.  Driving a car, understanding social networks, finding a good date, recommend a movie to watch, …
- Identify a specific task that can be done with machine learning
  - What is the objective of the task?
  - What is the data you need?
  - What type of ML problem is this?  Classification, regression, …
  - How would your approach compare to an expert-driven method?