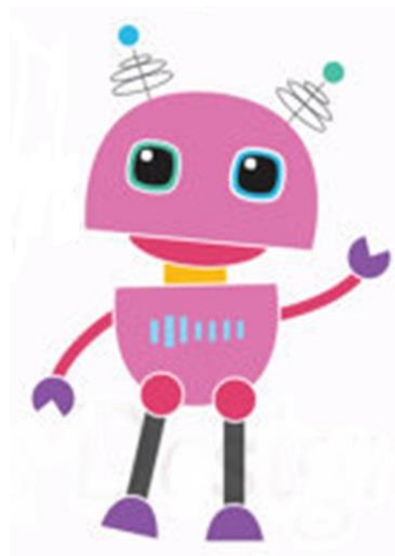
Machine Learning Introduction

Kate Saenko

Today

- Why do we need Machine Learning?
- What is Machine Learning?
- Topics we will cover
- Video "reading", quizzes, psets, etc.

Please, ask questions!



Introduction:
Why Do We Need ML?

Handwriting Recognition

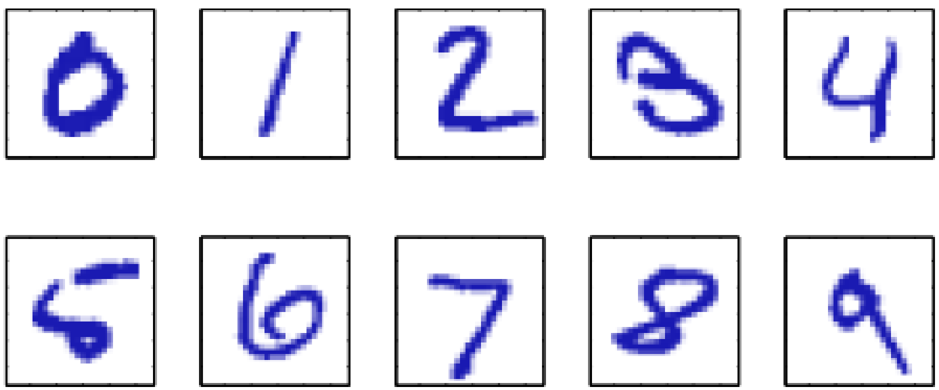


Figure from J.Corso

Computational Finance

US Markets | Actives | Gainers | Losers | Widely held | Dow 30

Jan 8 3:31pm †	Change	%Change	Level
Dow	-70.96	-0.81%	8,698.74
NASDAQ	+8.14	+0.51%	1,607.20
S&P	-2.09	-0.23%	904.56
DJ Wilshire 5000	-10.12	-0.11%	9,141.50
Russell 2000	+1.75	+0.35%	498.85
Philadelphia Semiconductor	-2.19	-0.98%	221.24
Dow Transports	-21.24	-0.60%	3,546.02
Dow Utilities	-1.53	-0.41%	370.78
NYSE Composite	+7.71	+0.13%	5,806.76
AMEX Composite	+35.10	+2.47%	1,456.56
Morningstar Index	-3.23	-0.15%	2,207.42
*10yr Note	-0.4900	-0.196%	2.445%
*NYMEX Crude Oil	-0.93	-2.18%	41.70
Gold	+12.80	+1.52%	854.50

Open Market Closed Pre-Market *as of previous close

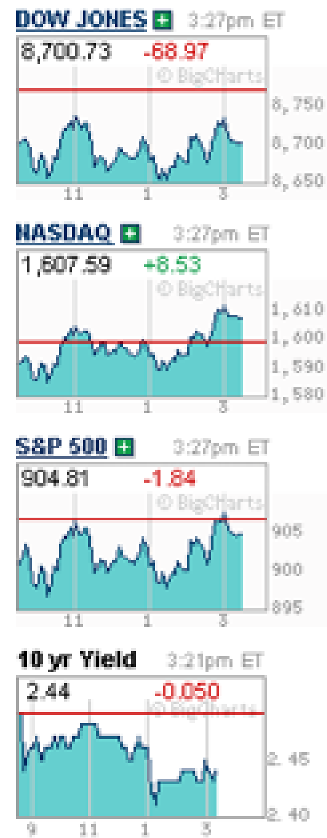
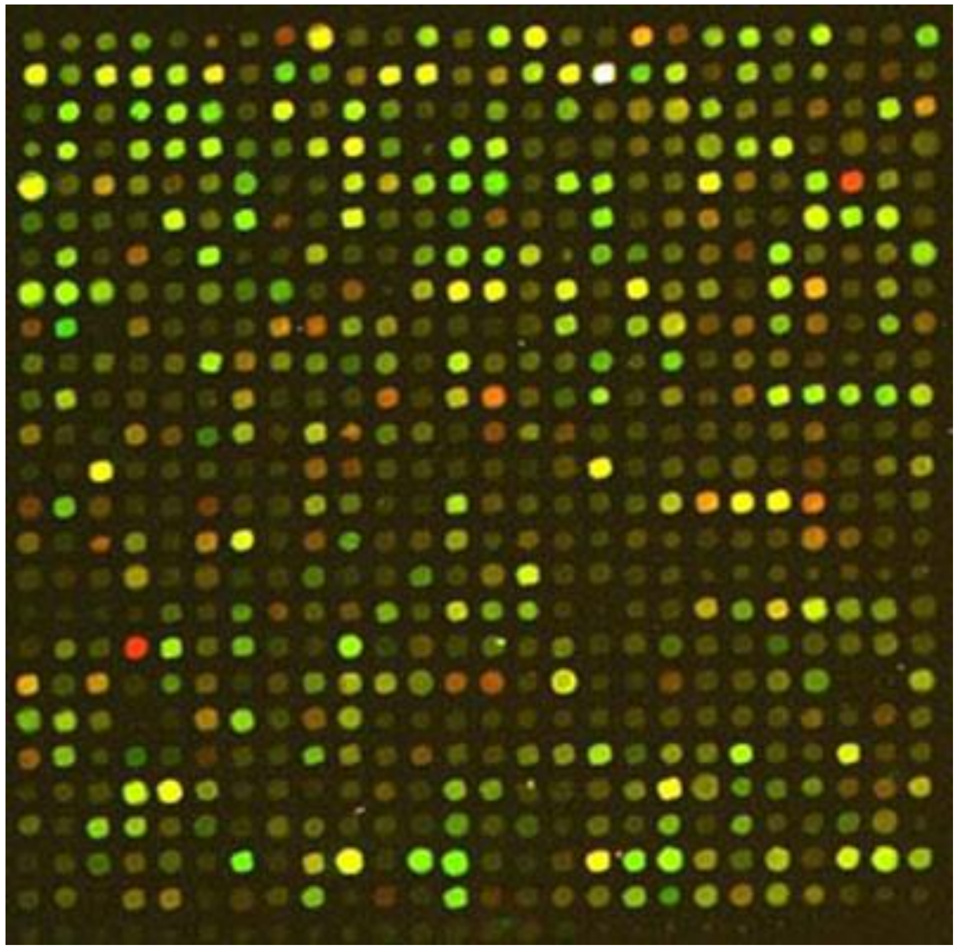


Figure from J.Corso

Gene Expression Analysis



Biometrics

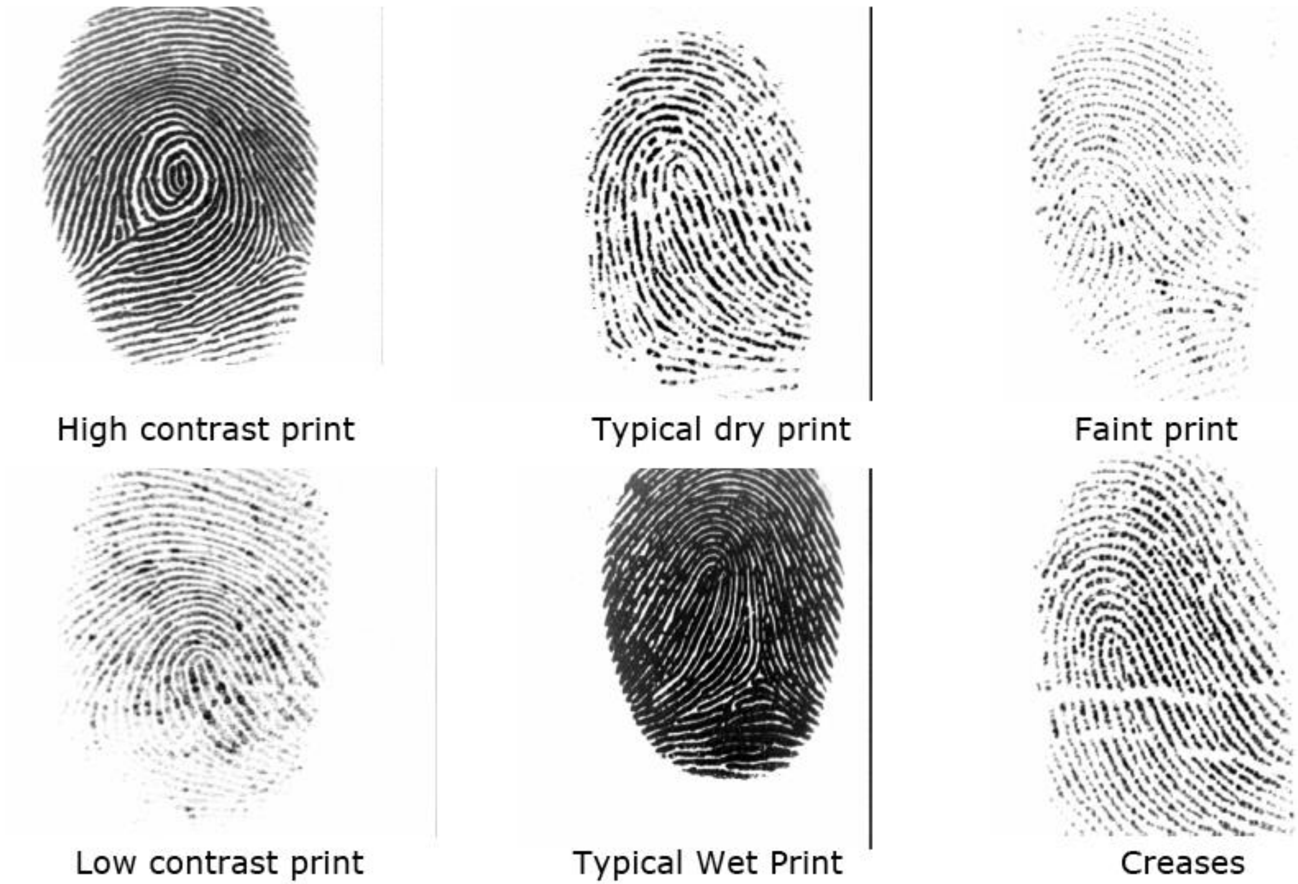


Figure from J.Corso

Crime Prediction?

- Person of Interest (TV Show, 2011-)

  A billionaire software-genius named Harold Finch creates a Machine for the government that is designed to detect acts of terror before they can happen, by monitoring the entire world by monitoring every cell-phone, email and surveillance camera.

--IMDb

Crime Prediction?

Introduction:
What is Machine Learning?

Machine Learning

- Branch of Artificial Intelligence
- *“creating machine algorithms that can learn from data”*
- Closely related to
  - Pattern recognition
  - Data Mining
  - Big Data
  - Deep learning

Example: Sorting Fish

Salmon

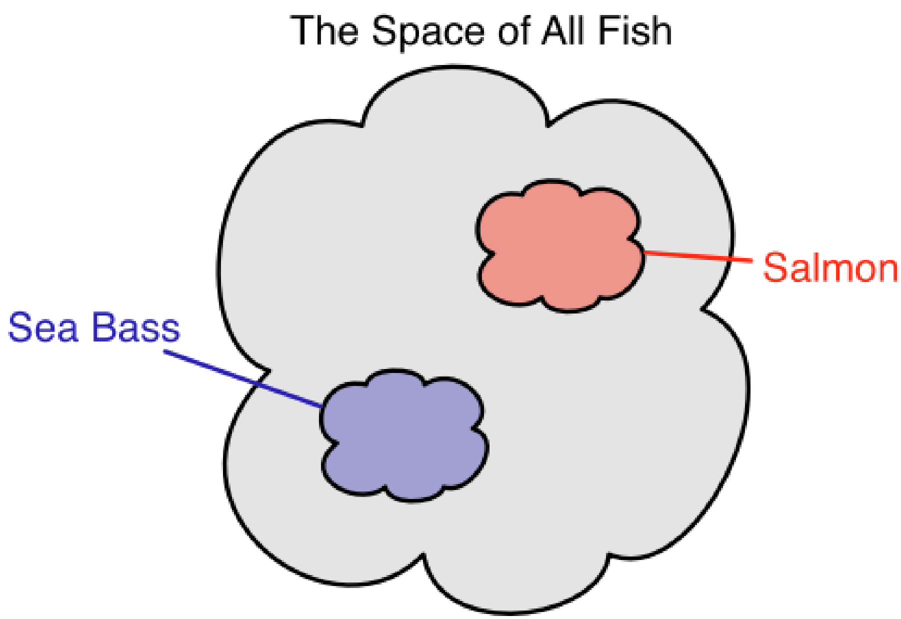
Sea Bass

Sorting Fish

- Set up a camera to measure the fish coming through on the conveyor belt (input)
- Classify each fish as salmon or sea bass (output)
- Prefer to mistake sea bass for salmon (cost function)

Salmon Sea bass

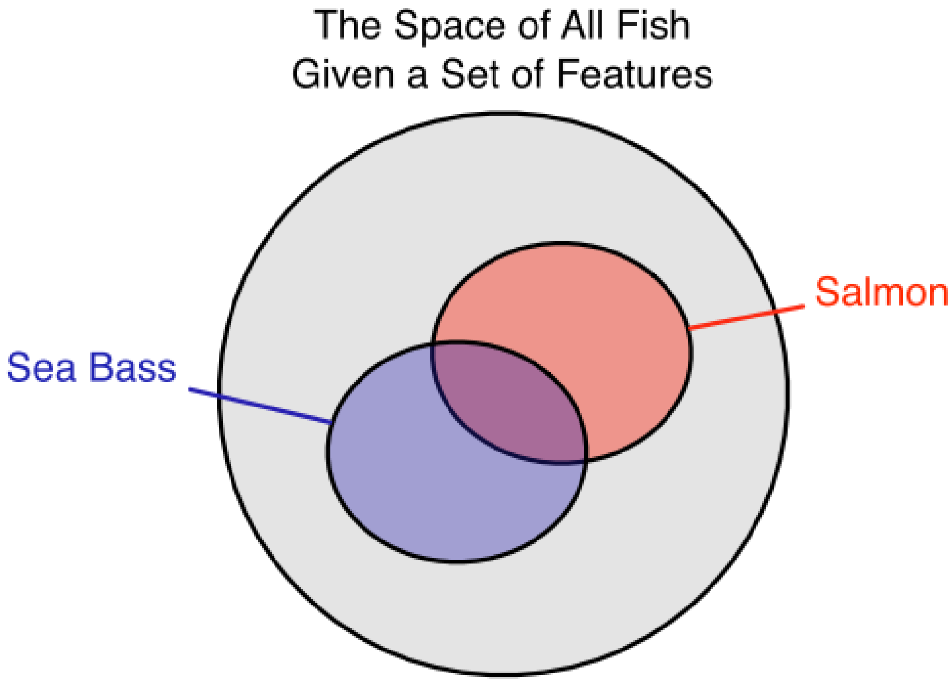
Ideal Feature Space



- Each dimension in the feature space is defined by some property of the fish – classes are separable
- Cannot measure most of them with the camera!

Figure from J.Corso

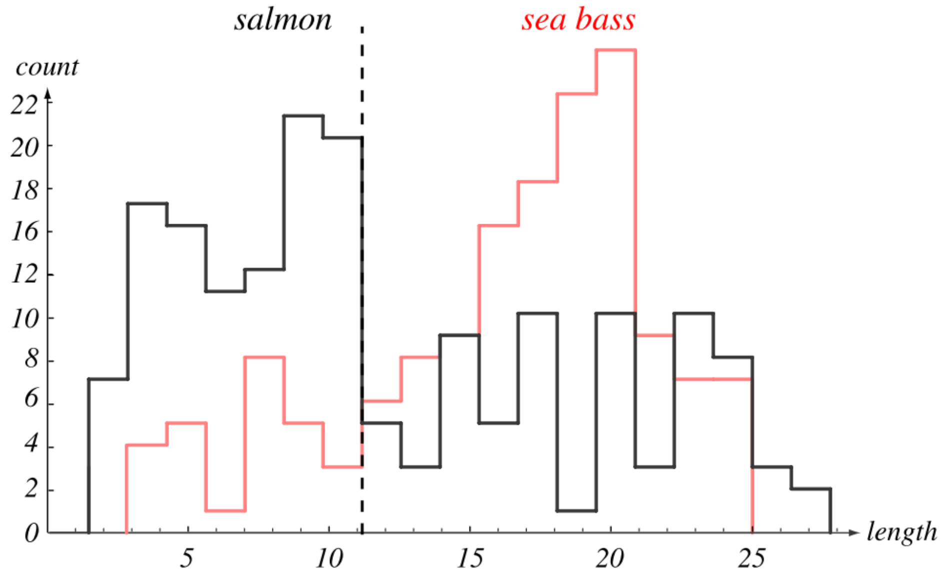
Real Feature Space



- In reality, we are given a set of observable features
- Project this very high dimension space down into a lower dimension – classes no longer separable!

Figure from J.Corso

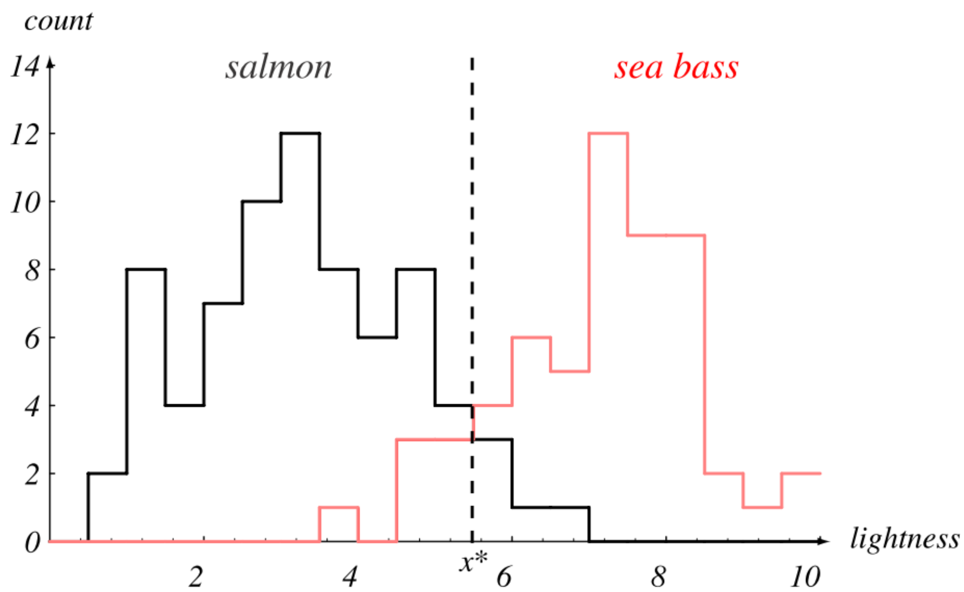
Length Feature



- Gather examples of length feature and plot its marginal distribution for each class
- Learn rule: if length > 11 inches, output “sea bass”
- Good enough?

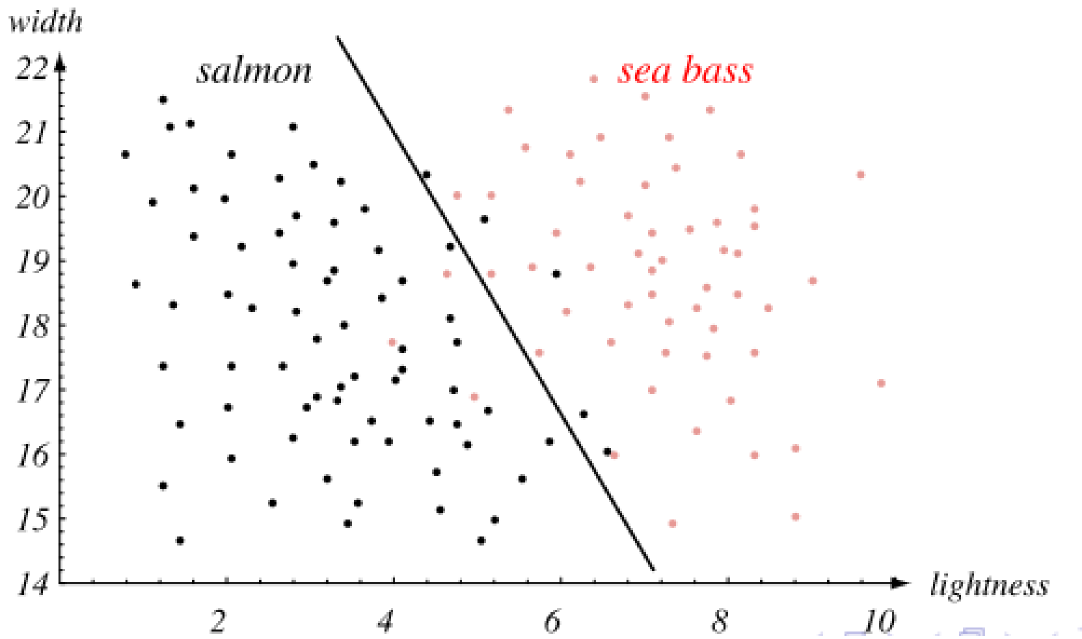
Figure from J.Corso

Try another feature: lightness



- Much better separation between the two classes
- Good enough?
- Single feature seldom sufficient!

2D Feature Space



- Combine lightness, x_1 and width, x_2

  $\mathbf{x} = [x_{1,}\, x_2\ ]^T$

- Plot joint distribution

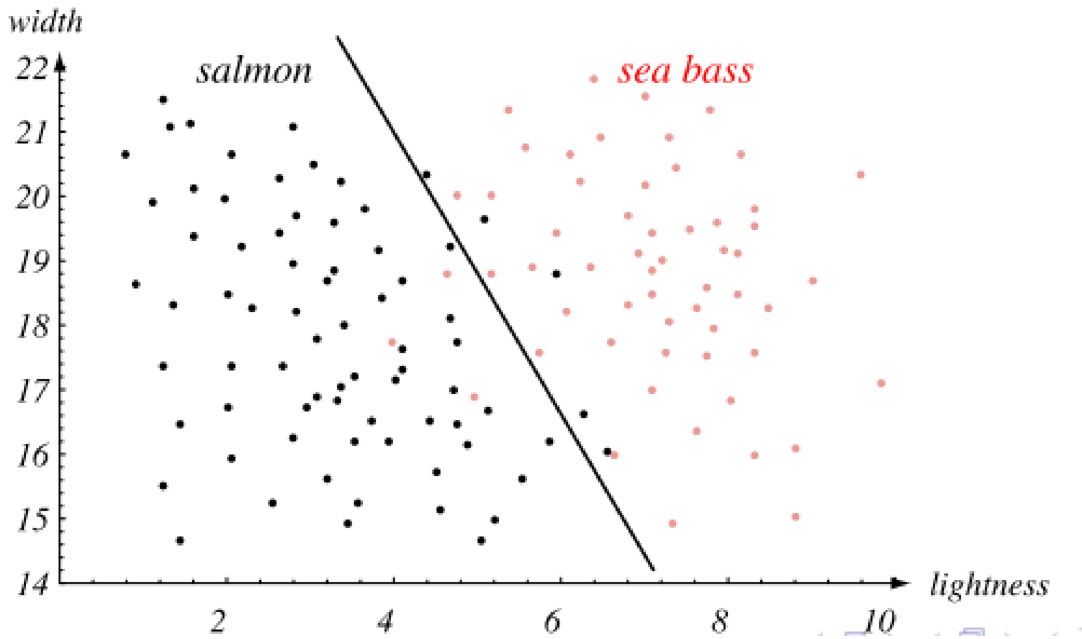
- Add more features?

Figure from J.Corso

Key questions

- How many features are required?
- Is there a point where we have too many features (curse of dimensionality)?
- How do we know beforehand which features will work best (feature selection)?
- Can we design an algorithm to learn features (representation learning)?
- What happens when there is feature redundance/correlation (dimensionality reduction)?

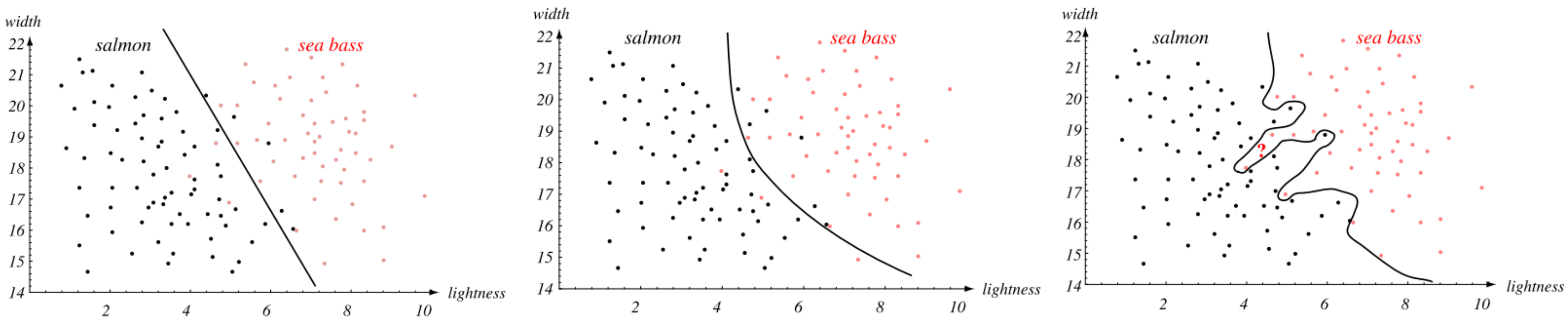
Decision Boundary



- The decision boundary is the sub-space in which classification among possible outcomes is equal
- Off the decision boundary, the decision is unambiguous

Figure from J.Corso

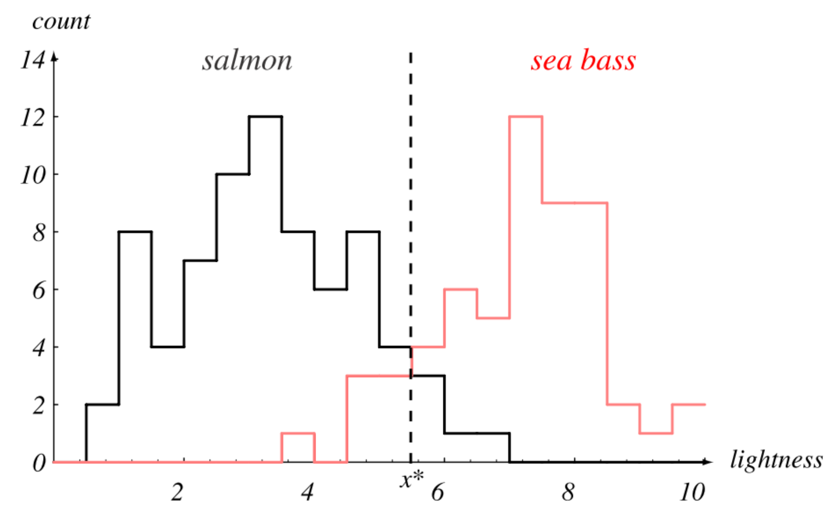
Bias-Variance Tradeoff



- Simple decision boundaries (e.g., linear) seem to miss some obvious trends in the data (bias)
- Complex decision boundaries seem to lock onto the idiosyncracies of the training data set (variance)
- Want classifiers that will work well on novel query data (generalization)

Decision Theory

- Wrong classifications are not always equally costly
- Recall: acceptable to have tasty pieces of salmon in cans labeled sea bass, but not otherwise



- Need to adjust our decisions (decision boundaries) to incorporate these varying costs
- How should we move the boundary for lightness?

Types of learning

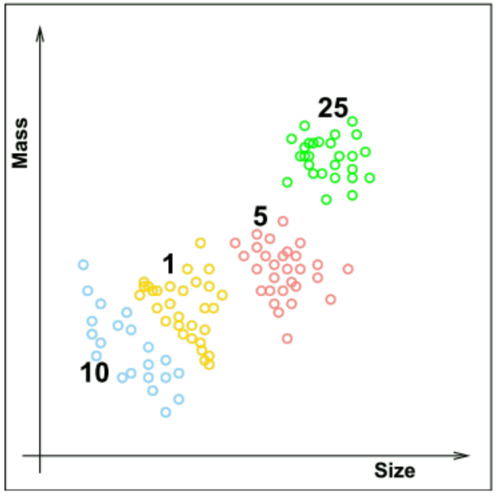
- Supervised
- Unsupervised
- Reinforcement

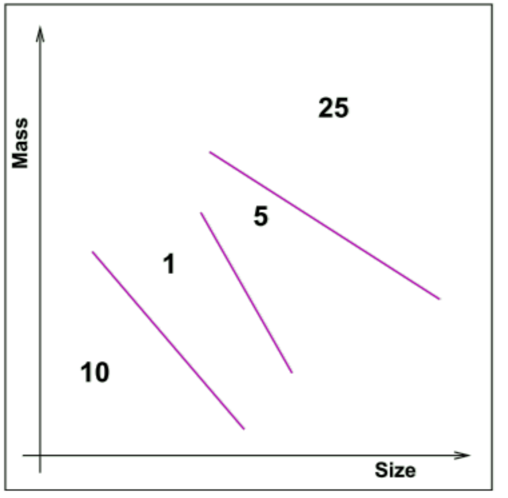
Supervised Learning

- Given a training set consisting of inputs and outputs, learn to map novel inputs to outputs
- The novel inputs are called a test set
- Outputs can be
  - Categorical (classification)
  - Continuous (regression)

Example of Supervised Learning

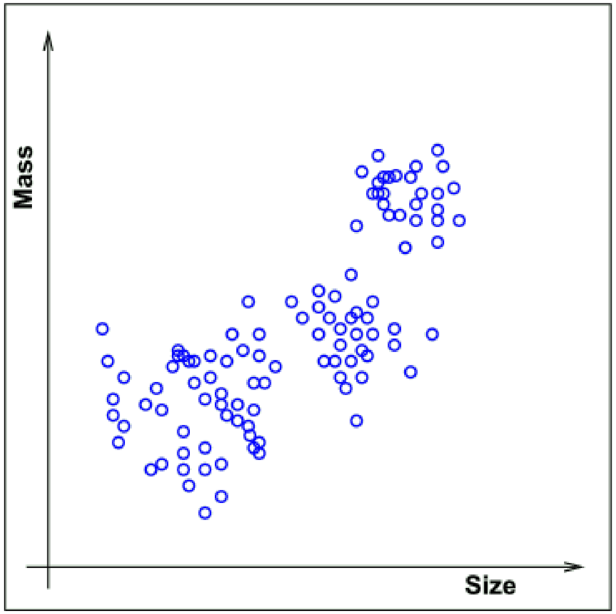
coin classification





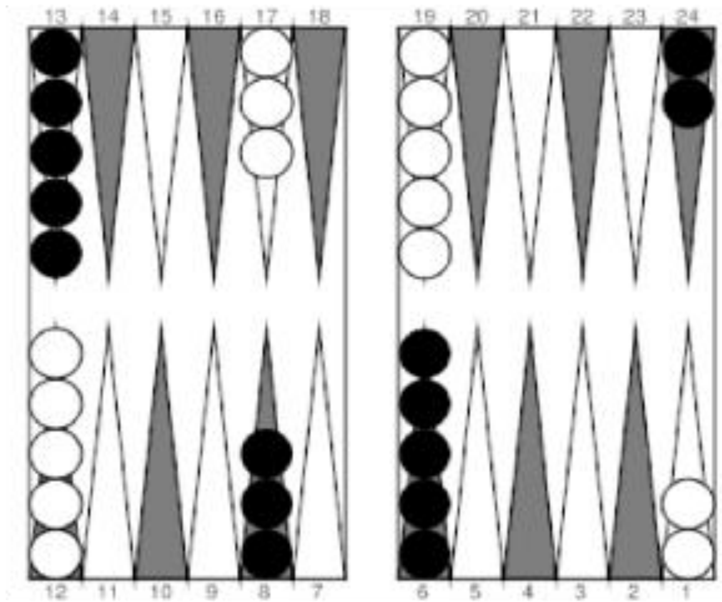
- Given training set consisting of coin denomination (penny, nickel, dime, quarter), mass and size
- Learn to predict denomination
- What is input? Output?

Unsupervised Learning



- Given training set consisting of ~~coin denomination (penny, nickel, dime, quarter)~~ mass and size
- Learn… something?

Reinforcement Learning



- Given only input
- Predict output, get a grade for it

Types of learning

1. Suppose you are building a Machine to detect people that are about to commit a crime by analyzing camera surveillance and phone conversations. Would you model it as a classification or a regression problem?
2. You are designing a robot that can teach itself to play chess by playing lots of games with you and your friends. Would you use supervised, unsupervised, or reinforcement learning?

Machine Learning

Introduction:
Course Overview

Topics

- Linear Regression
- Bias/Variance Tradeoff
- Bayesian Regression

Topics

- Logistic Regression
- Regularization
- Neural Networks
- Support Vector Machines

Topics

- Ensemble Classifiers, Boosting
- Bayesian Classifiers, MAP
- Nearest Neighbor Methods
- Metric Learning

Topics

- Unsupervised Learning
- Clustering
- Dimensionality Reduction
- Anomaly Detection

Logistics

Course number: 91.545 (graduate) 91.422(undergraduate)

Instructor: Kate Saenko, saenko@cs.uml.edu

TA: Huijuan Xu

Location: Olsen Hall, Room 102

Meeting Times: TR 4:00-5:15pm

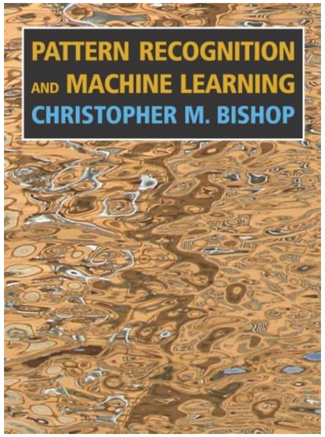
Webpage:
https://https://sites.google.com/site/umlcsmachinelearningspring2015//

Office Hours:

Kate: TBD Olsen 223

Huijuan: TBD

Textbook



- Required textbook

Bishop, C. M. Pattern Recognition and Machine Learning. Springer. 2007

- Other textbooks

Duda, R.O., Hart, P.E., and Stork, D.G. Pattern Classification. Wiley-Interscience. 2nd Edition. 2001.

Marsland, S. Machine Learning: An Algorithmic Perspective. CRC Press. 2009.Theodoridis, S. and Koutroumbas, K. Pattern Recognition. Edition 4. Academic Press, 2008.

Russell, S. and Norvig, N. Artificial Intelligence: A Modern Approach. Prentice Hall Series in Artificial Intelligence. 2003.

Bishop, C. M. Neural Networks for Pattern Recognition. Oxford University Press. 1995.

Hastie, T., Tibshirani, R. and Friedman, J. The Elements of StatisticalLearning. Springer. 2001.

Koller, D. and Friedman, N. Probabilistic Graphical Models. MIT Press. 2009.

Video “textbook”

- We will closely follow Stanford’s Machine Learning course video lectures

  https://class.coursera.org/ml-003/lecture/preview

  How it works
  - Each week, watch ~1hr video before class
  - Be prepared to discuss in class
  - I will re-cap and introduce more in-depth and additional material

Grading

- Four contributions to grade
  - best 5 of 6 homeworks: 25% (5% each)
  - 3 quizzes: 45% (Quiz0 5%, Quiz1&2 40%)
  - final project: 25%,
  - class participation: 5%

Problem Sets

- Octave
  - scientific programming language
  - free version of Matlab
- 6 assignments
  - roughly every two weeks
  - should take 5-10 hours each
  - prepare you for the quizzes!
- Self-grading
  - you will submit code, answers, and your own grade
  - we will randomly run code to verify

Late policy

- 20% off per day
- Up to 4 days

Final Project

- Novel implementation and evaluation of a machine learning algorithm
- Apply ML to a problem that interests you, using a publicly available dataset if possible
- Teams of 3 required
- Four components
  - Proposal/pitch
  - Implementation
  - Evaluation
  - Final write up (conference paper style)

More later…

Example Projects

Predicting Movie and TV Preferences from Facebook Profiles

Classifying events in videos, e.g. birthday party

Speech Recognition: build Siri from scratch

Associating enhancers to the genes they regulate

Predicting the future with deep learning

…

More ideas here

http://cs229.stanford.edu/projectIdeas_2012.html

First Homework, for Thursday

- Watch the video lectures

I. Introduction (Week 1)

Welcome (7 min)

What is Machine Learning? (7 min)

Supervised Learning (12 min)

Unsupervised Learning (14 min)

III. Linear Algebra Review (Week 1, ~~Optional~~)

Matrices and Vectors (9 min)

Addition and Scalar Multiplication (7 min)

Matrix Vector Multiplication (14 min)

Matrix Matrix Multiplication (11 min)

Matrix Multiplication Properties (9 min)

Inverse and Transpose (11 min)

Quiz 0

- due Thu Jan 22 4:00pm (in class)
- 5% of grade
- Self-examination of the math prerequisite knowledge
  - Probability & statistics, discrete math, calculus, basic matrix algebra
- Should be able to do >60% to proceed in the course