Welcome to CMPS 142 Machine Learning

- Instructor: David Helmbold, dph@soe.ucsc.edu
  - Office hours: Tentatively after class – Wed-Fri 3:30-4:30.
- TA: Michael Shavlovsky, mshavlov@ucsc.edu
- Web page:

  https://cmps142-winter16-01.courses.soe.ucsc.edu/

  Also Piazza discussion forum
- Webcast info: Account: cmps142 password: mlvideo
- Text*:* Andrew Ng's lecture notes*:*

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

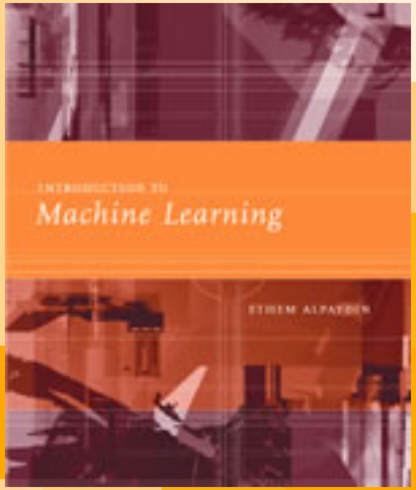
Administrivia

- Sign up sheet (enrollment)
- Evaluation:
  - Group Homework ~50%
  - Final exam ~50%

  Must pass exam
- Expectations/Style
  - Reading assignments
  - Attendance/participation
  - My hearing/writing
  - Academic honesty

Topics:

- Introduction
- Regression and multiclass (ch 3)
- Logistic regression
- Probability review
- Naïve Bayes and generative models
- Perceptron Algorithm
- Support Vector Machines
- Decision trees
- Neural Networks
- Model and feature selection
- Ensemble methods
- Learning Theory
- Unsupervised learning



Lecture Slides for

INTRODUCTION TO

Machine Learning

ETHEM ALPAYDIN

(modified by DPH 2006--2011)
alpaydin@boun.edu.tr
http://www.cmpe.boun.edu.tr/~ethem/i2ml

CHAPTER 1:

Introduction

Why "Learn"?

- Machine learning is programming computers to optimize a performance criterion using example data or past experience (<u>inference</u> in statistics)
- 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, object detection)
  - Solution changes in time (routing on a computer network)
  - Solution needs to be adapted or customized to particular cases (or users)

What We Talk About When We Talk About "Learning"

- Learning general models from a set 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 "Da Vinci Code" also bought "The Five People You Meet in Heaven" (www.amazon.com)

- Build a model that is *a good and useful approximation* to the data.

What is Machine Learning?

- 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

Statistical Machine learning is not:

- Cognitive science (how people think/learn)
- Teaching computers to think

But is related to:

- Statistics
- Data Mining – Knowledge Discovery
- Control theory
- part of AI, but not “traditional” AI

Supervised Batch Learning

- Assume (unknown) distribution over "things"
- Things have measurable ***attributes*** or ***features***
- Get ***instances*** (feature vectors) ***x*** by drawing things from distribution and recording observations.
- Teacher ***labels*** instances making ***examples*** (***x***, *y*)
- Set of labeled examples is the ***training set*** or ***sample***
- Create ***hypothesis*** (rule or function) from sample
- hypothesis predicts on *new* random instances, evaluated using a ***loss function*** (e.g. number of mistakes)

Supervised Learning (cont.)

- **Classification**: labels are nominal (unordered set, e.g. {ham, spam} {democrat, republican, indep.})
- Binary Classification
- **Regression**: labels are numeric (e.g. price of house)
- **Ranking** problems (order a set of objects)

Examples

Thing	Observations	Prediction
Written Digit	Pixel array	Which digit? classification
Email message	Words, Subject, sender	Ham or Spam? binary classification
Customer	Recent purchases	interest level in a new product probability
Used car	Year, make, mpg, options	Price or value regression

Batch Assumption: iid Examples

- Distribution of things and measurements defines some unknown (but fixed) $P(\boldsymbol{x},y)$ or $D(\boldsymbol{x},y)$ over domain-label pairs
- Find a hypothesis or function $f(\boldsymbol{x})$ that is "close" to the "truth"
  - A loss function $L(y, y')$ measures error of predictions, often $L(y,y')=0$ if $y=y'$ and $L(y,y')=1$ otherwise (classification)
  - Want to minimize $\int P(\boldsymbol{x},y)\, L(y, f(\boldsymbol{x}))$

    -- e.g. probability of error for 0-1 loss

Supervised Learning: Uses

- 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 and data entry errors

Can we Generalize?

- Learning is an ill-posed problem:

  If we assume nothing else, any label y could be right for an unseen $\boldsymbol{x}$
- Need an ***inductive bias*** limiting possible P($\boldsymbol{x}$,y)
- Often assume some kind of simplicity (e.g. linearity) based on domain knowledge
- Bayesian approach: put *prior* on rules, and balance prior with evidence (data)

Noise

Data not always perfect

- Unmeasured Features
- Attribute noise (random or systemic)
- Label noise (random or systemic)

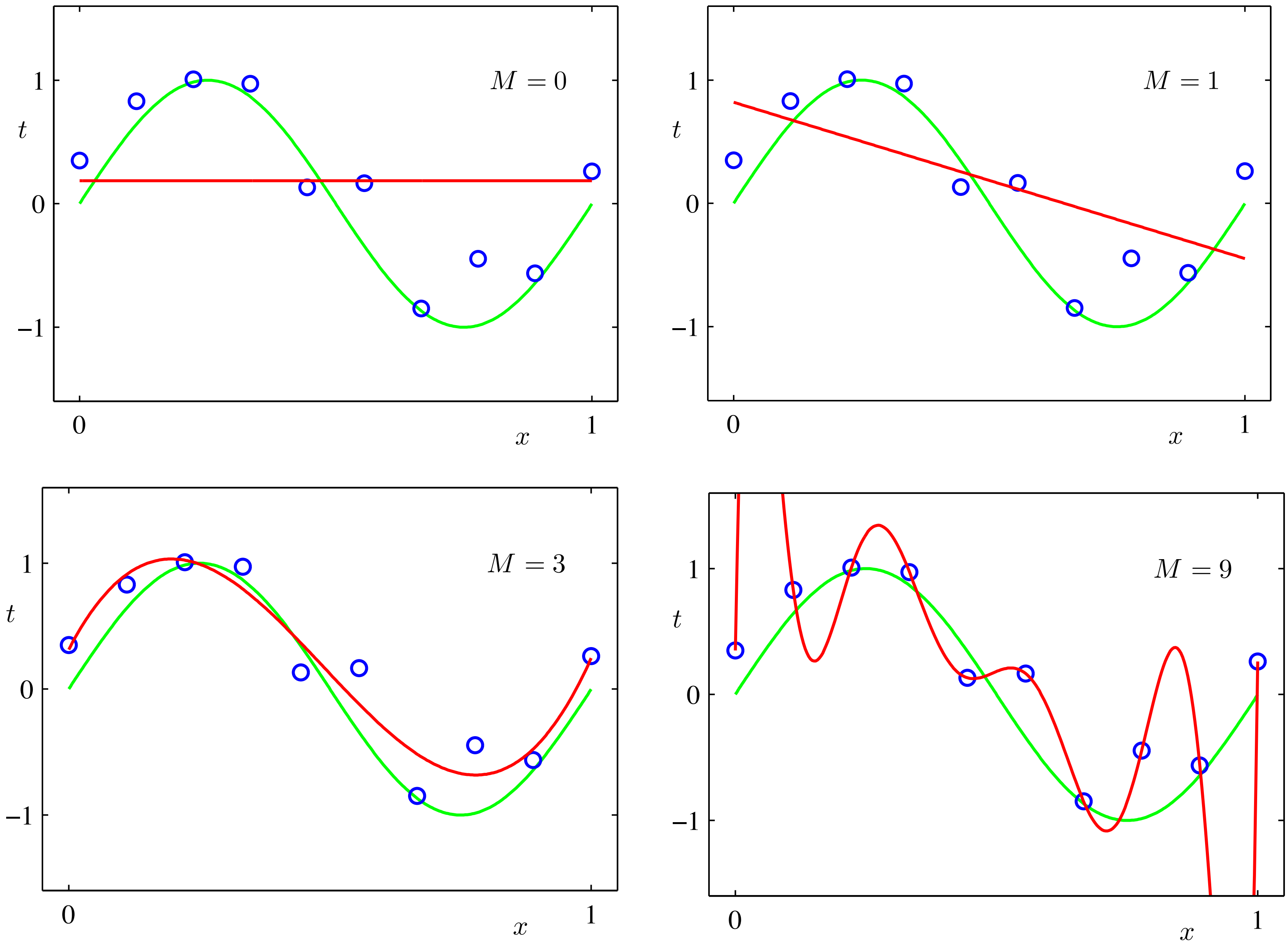
- inductive bias errors may look like noise

Overfitting and Underfitting

- **Overfitting** happens when the hypothesis is too complex for the "truth"

- **Underfitting** happens when the hypothesis is too simple.

Bishop fig 1.4



Don't rely on training error!

- To estimate generalization error, we need data **unseen** during training. Often data split into
  - Training set (70%)
  - Validation set (10%) (did training work? Use for Parameter selection/model complexity)
  - Final Test (publication) set (20%)
- Resampling when there are few examples

  cross validation (describe)

Other kinds of supervised learning

- Reinforcement learning - learning a policy for influencing or reacting to environment
  - Game playing/robot in a maze, etc.
  - No supervised output, but delayed rewards
  - Credit assignment problem

- On-line learning: predict on each instance in turn

- Semi-supervised learning uses both labeled and unlabeled data

- Active learning – request labels for particular instances

Unsupervised Learning

- Learning "what normally happens"
- No labels
- Clustering: Grouping similar instances
- Example applications
  - Segmentation in customer relationship mgmt
  - Image compression: Color quantization
  - Bioinformatics: Learning motifs
  - Identifying unusual Airplane landings
  - Deep learning – learn the "features"