Overview

COMS 4771 Fall 2019

Overview

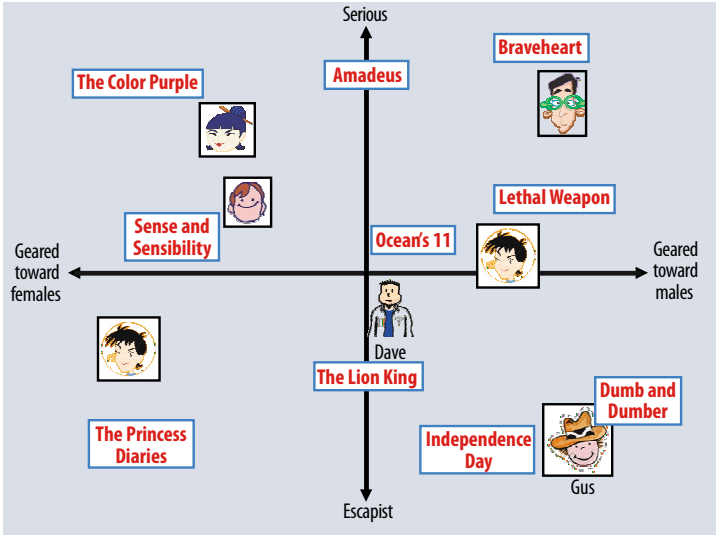
- What is machine learning?
- Basic topics/challenges in machine learning

Applications I

- Image classification: Predict bird species depicted in image

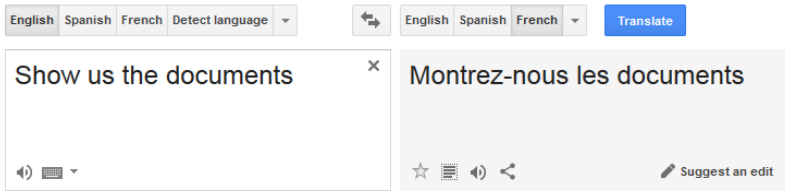
Applications II

- Recommender systems: Predict how user would rate a movie (Koren, Bell, and Volinsky, 2009)



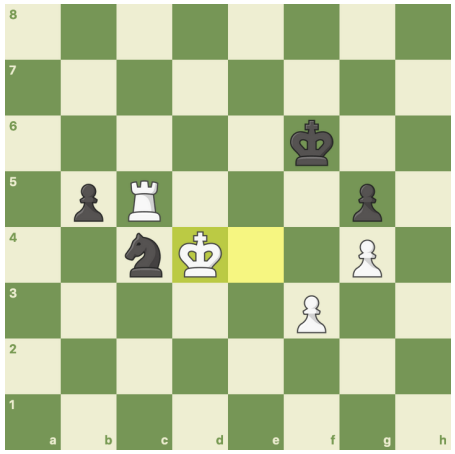
Applications III

- Machine translation: Predict French translation of English sentence (Google translate)



Applications IV

- Chess: Predict win probability of a move in given game state (AlphaZero)



Work in ML

- Applied ML
  - Collect/prepare data, build/train models, analyze errors
- ML developer
  - Implement ML algorithms and infrastructure
- ML research
  - Design/analyze models and algorithms

Mathematical and computational prerequisites

- Math
  - Linear algebra, probability, multivariable calculus, reading and writing proofs
- Software/programming
  - Much ML work is implemented in python with libraries such as numpy and pytorch

Basic setting: supervised learning

- Training data: dataset comprised of *labeled examples*
  - *Labeled example*: a pair (input, label)
- Goal: learn function to predict label from input for new examples

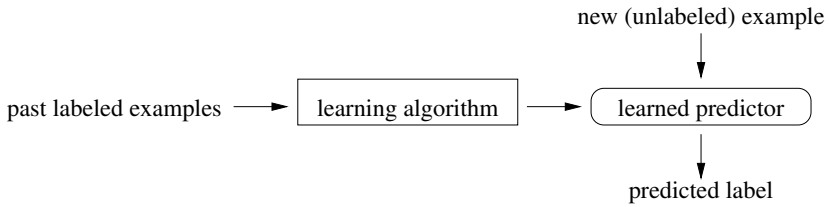


Figure 1: Schematic for supervised learning

Examples of functions I

- Decision tree

if age ≥ 40 **then**
  if genre $=$ western **then**
    return 4.3
  else if release date > 1998 **then**
    return 2.5
  else
    ⋮
  end if
else if $\cdots$ **then**
  ⋮
end if

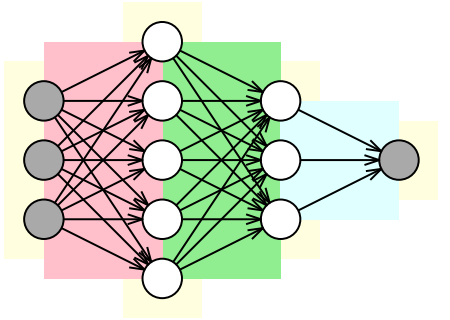
Examples of functions II

- Linear classifier

if $0.335 \cdot x_1 + 2.5 \cdot x_2 + \cdots + 6.35 \cdot x_{10^6} > 4.3$ **then**
  return spam
else
  return not spam
end if

Examples of functions III

- Neural network



input hidden units output

Types of prediction problems

- Binary classification
  - Given an email, is it spam or not?
  - (What's the probability that it is spam?)
- Multi-class classification
  - Given an image, what animal is depicted?
  - (Or which animals are depicted?)
- Regression
  - Given clincal measurements, what is level of tumor antigens?
  - (In absolute level? Log-scale?)
- Structured output prediction
  - Given a sentence, what is its grammatical parse tree?
  - (Or dependency tree?)
- ...

Beyond supervised learning

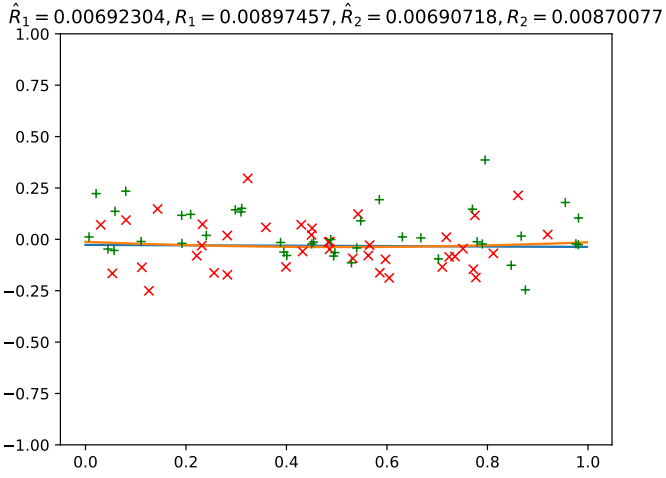
- Unsupervised learning / probabilistic modeling
- Online learning
- Reinforcement learning

Challenges in supervised learning

- Might not have the right data
- Might pick a bad model
- Might not fit training data well (under-fitting)
- Might fit the training data too well (over-fitting)
- Training data could be noisy / corrupted (robustness)
- ...

Example: over-fitting

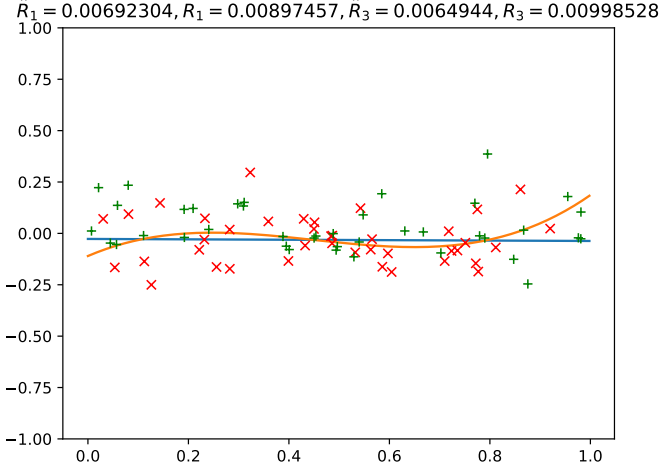
- Which polynomial degree to use?
- Truth: $y = 0 \cdot x +$ noise
- Red points: training data
- Green points: unseen data



Example: over-fitting

- Which polynomial degree to use?
- Truth: $y = 0 \cdot x +$ noise
- Red points: training data
- Green points: unseen data

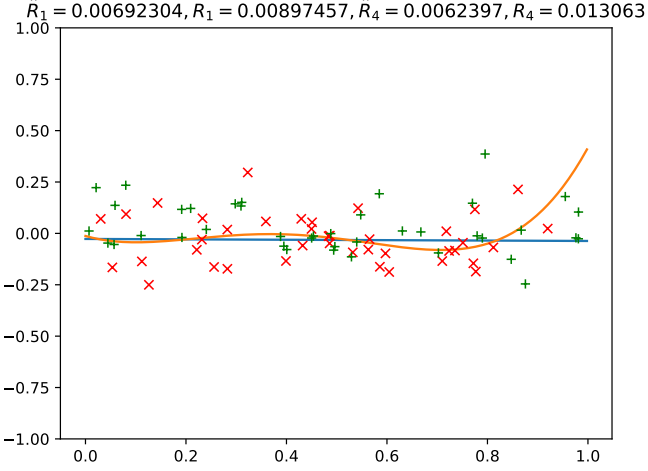
$\hat{R}_1 = 0.00692304, R_1 = 0.00897457, \hat{R}_3 = 0.0064944, R_3 = 0.00998528$



Example: over-fitting

- Which polynomial degree to use?
- Truth: $y = 0 \cdot x +$ noise
- Red points: training data
- Green points: unseen data

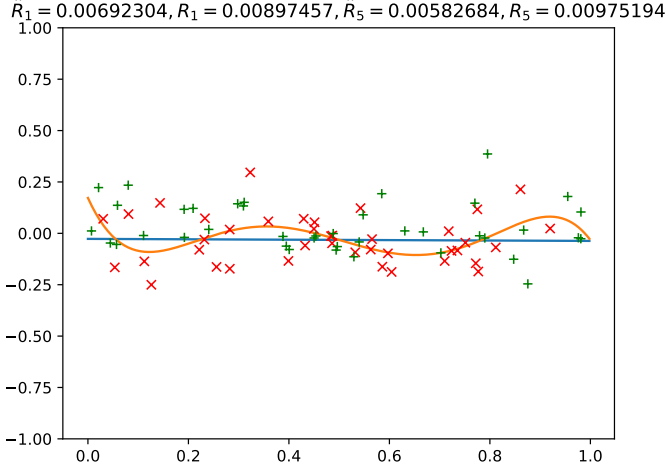
$\hat{R}_1 = 0.00692304, R_1 = 0.00897457, \hat{R}_4 = 0.0062397, R_4 = 0.013063$



Example: over-fitting

- Which polynomial degree to use?
- Truth: $y = 0 \cdot x +$ noise
- Red points: training data
- Green points: unseen data

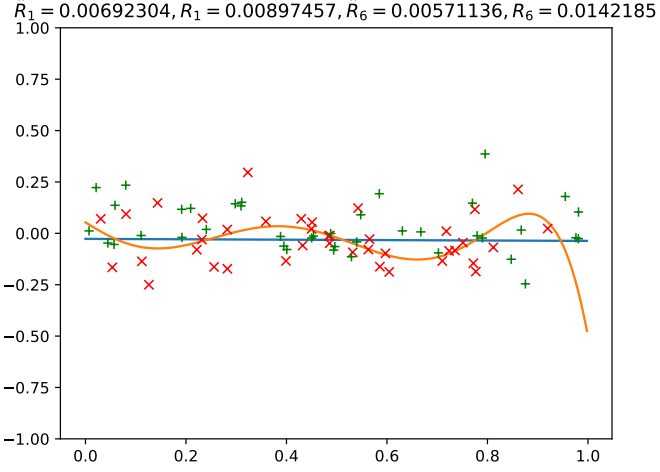
$\hat{R}_1 = 0.00692304, R_1 = 0.00897457, \hat{R}_5 = 0.00582684, R_5 = 0.00975194$



Example: over-fitting

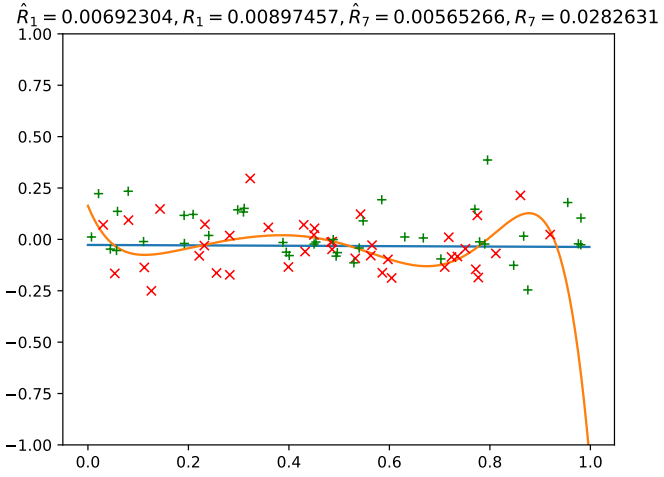
- Which polynomial degree to use?
- Truth: $y = 0 \cdot x +$ noise
- Red points: training data
- Green points: unseen data

$\hat{R}_1 = 0.00692304, R_1 = 0.00897457, \hat{R}_6 = 0.00571136, R_6 = 0.0142185$



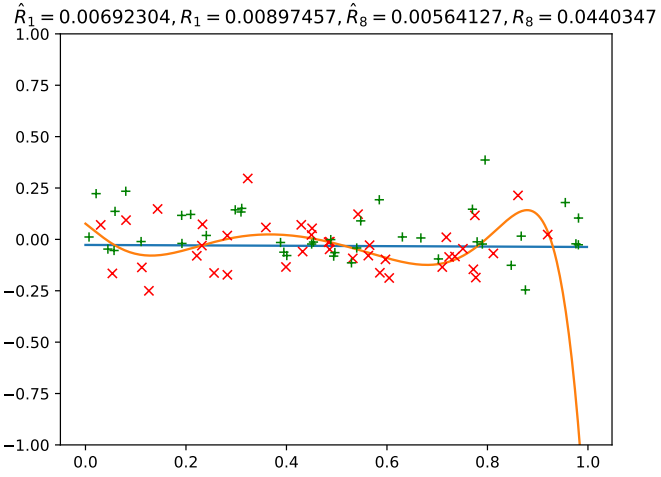
Example: over-fitting

- Which polynomial degree to use?
- Truth: $y = 0 \cdot x +$ noise
- Red points: training data
- Green points: unseen data



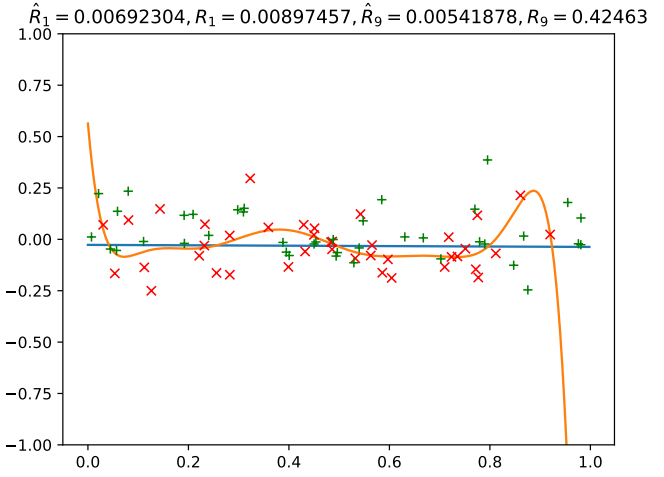
Example: over-fitting

- Which polynomial degree to use?
- Truth: $y = 0 \cdot x +$ noise
- Red points: training data
- Green points: unseen data



Example: over-fitting

- Which polynomial degree to use?
- Truth: $y = 0 \cdot x +$ noise
- Red points: training data
- Green points: unseen data



Example: the right data

- Given a college applicant, will they graduate if admitted?
- What is appropriate training data?
  - input = past applicant; label = admitted or not
  - input = past admit; label = graduated or not
  - input = past applicant; label = graduated or not

Overview of the rest of the course

- Non-parametric methods
  - Simple and flexible methods for prediction
- Prediction theory
  - Statistical model for studying prediction problems
- Regression
  - Models and methods for predicting real-valued outcomes
  - Inductive bias, features, kernels
- Classification
  - Models and methods for predicting discrete-valued outcomes
  - Surrogate losses, margins, cost-sensitive risk, fairness, ensemble methods
- Optimization
  - Convex optimization and neural network training
- Unsupervised learning
  - Methods for clustering and matrix approximation