All models are wrong, some models are useful.

- George Box, Statistician

CSE455 & CSE 552
Machine Learning

Spring 2025 Semester

Introduction

1

Today

- Introductions
- Administrative Details
- Syllabus
- Introduction to ML
- Linear Classifiers

2

Introductions

- Dr. Genc – Instructor:
  - PhD from CS @ UIUC in 1999
  - 20+ years research and R&D management experience
  - Tel: x2220
- Office Hours
  - Mondays 3:30pm-4:30pm

3

Admin: Communication

- The course communication will be done via Teams
  - CSE552 ML 2025 Spring | General | Microsoft Teams
  - Make sure you register as soon as possible

4

Admin: Grading

- Grading for undergrads and grads
  - %35 – Homeworks
  - %25 – Project
  - %20 – Midterm
  - %20 – Final

5

Admin: Reading Material

- Book (suggestions):
  - Introduction to Machine Learning by Ethem Alpaydin
  - Machine Learning by Tom M. Mitchell
- Other reading materials will be provided as needed

6

Admin: Assignments

- Homework
  - Implementations (5-6)
  - Python programming language
  - Datasets will be provided
  - No late submission

7

Admin: Projects & Presentations

- Projects
  - A problem and solution
  - List to be provided in two weeks
  - Presentations and demos to be done at the end of the term

8

Aim of this Course

- Lectures to discuss basics of ML methods
- In depth discussions on the implications on real world problems
  - Lots of buzz about ML and its application – let's get on the bandwagon!
  - Big data …
  - AI
  - Superintelligence
- Let's learn through trying on real problems…

February 2025 CSE455/CSE552 Machine Learning 9

9

Human vs Artificial Intelligence

February 2025 CSE455/CSE552 Machine Learning 10

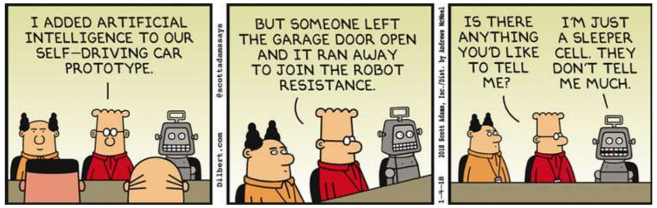
10

Why are you here?

February 2025 CSE455/CSE552 Machine Learning 11

11

Why are you here?



February 2025 CSE455/CSE552 Machine Learning 12

12

Introduction

Some of the following slides are adapted from E. Alpaydin

13

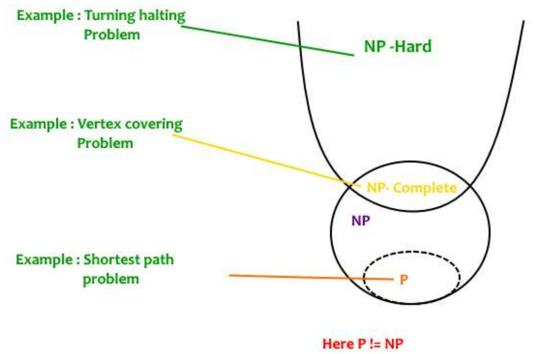
Why Machine Learning?

- Most of the real world problems are:
  - NP-Hard
    - scene matching, big data problems …
  - Ill-defined
    - 3D reconstruction from a single image, missing data
  - The right answer is subjective
    - Image segmentation, language, …
  - Hard to model
    - scene classification, customer behavior
- Machine Learning tries to use statistical reasoning to find **approximate** solutions for tackling the above difficulties

Joshi et al

14

NP-Hard



15

Halting Problem

Can you write the Python procedure halts(s):

- Input:
  - s: a string representing a Python program.
- Output:
  - true: if evaluating the input program would ever finish.
  - false: otherwise.

16

Halting Problem

Proving non-existence by contradiction:
1. Show X does not make sense.
2. Show that if you have A you can make X.
3. Therefore, A must not exist.

```
def paradox():
    if halts(paradox):
        while True:
            pass
```


17

Halting Problem

Proving non-existence by contradiction:
1. Show X does not make sense.
2. Show that if you have A you can make X.
3. Therefore, A must not exist.

Noncomputability

(Alan Turing)

X = paradox procedure

A = Algorithm that solves HALTING problem

18

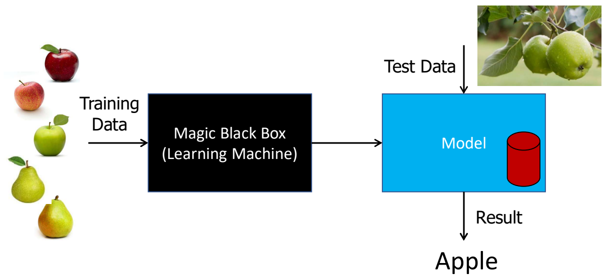
What is Machine Learning

- … a computer program that can learn from experience with respect to some class of tasks and performance measure …" – Mitchell 1997
- " … ML is the science of getting computers to learn, without explicitly programmed…" –Ng
- ML is more than just memorizing facts:
  - learning the underlying structure of the problem or data
- Also known as:
  - Regression
  - Pattern recognition
  - Data mining

Dönnes

19

Black Box Model of ML



Training data: Sample images of the class of objects to be learned

Model: The model can identify any koala images...

20

Learning Algorithms

- Supervised Learning
  - Generative/Discriminative models
  - ANN/Boosting/Decision Tree/NNA/Random Forests
- Unsupervised Learning
  - Clustering
  - K-Means/Dirichlet/Gaussian Processes/EM
- Semi-Supervised Learning
  - Constrained Clustering/Distance Metric Learning/Manifold based Learning/Compressed Sensing/Active Learning

Joshi et al

21

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)

22

When We Talk About "Learning" ...

- Learning general models from a data of particular examples
- Data is cheap and abundant (data warehouses, data marts); knowledge/model 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

23

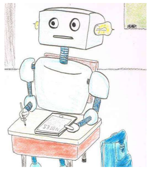
Data Mining

- **Retail**: Market basket analysis, Customer relationship management (CRM)
- **Finance**: Credit scoring, fraud detection
- **Manufacturing**: Optimization, troubleshooting
- **Medicine**: Medical diagnosis
- **Telecommunications**: Quality of service optimization
- **Bioinformatics**: Motifs, alignment
- **Web mining**: Search engines
- ...

24

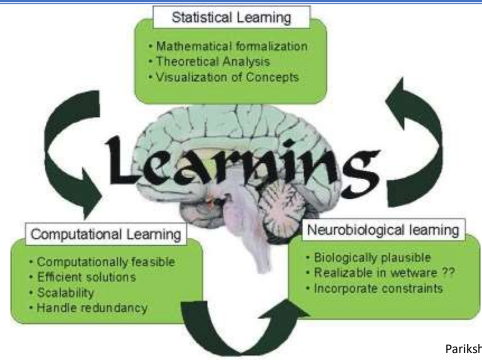
What is Machine Learning?

- 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



25

Artificial Intelligence



Parikshit Ram, GA Tech

26

Applications

- Association
- Supervised Learning
  - Classification
  - Regression
- Unsupervised Learning
- Reinforcement Learning
- …
- Sparse/Compressed Learning

27

Machine Learning – Another View

28

Learning Associations

- Basket analysis:

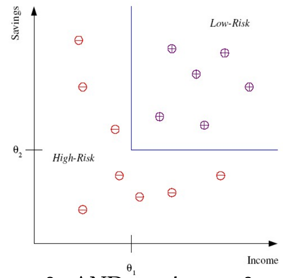
  P (*Y* | *X*) probability that somebody who buys *X* also buys *Y* where *X* and *Y* are products/services.

  Example: *P* (ayran | pide) = 0.7

29

Classification

- Example: Credit scoring
- Differentiating between low-risk and high-risk customers from their *income* and *savings*



Discriminant: IF *income* > θ_1 AND *savings* > θ_2 THEN low-risk ELSE high-risk

30

Classification: Applications

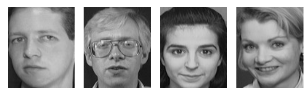
- Aka "Pattern Recognition"
- **Face recognition**: Pose, lighting, occlusion (glasses, beard), make-up, hair style
- **Character recognition**: Different handwriting styles
- **Speech recognition**: Temporal dependency.
  - Use of a dictionary or the syntax of the language.
  - Sensor fusion: Combine multiple modalities; eg, visual (lip image) and acoustic for speech
- **Medical diagnosis**: From symptoms to illnesses
- ...

31

Face Recognition

Training examples of a person

Test images

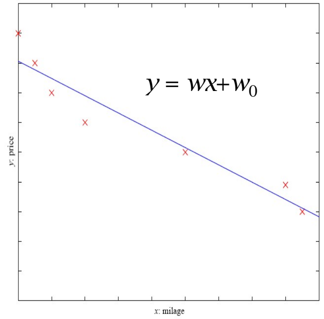


AT&T Laboratories, Cambridge UK
http://www.uk.research.att.com/facedatabase.html

32

Regression

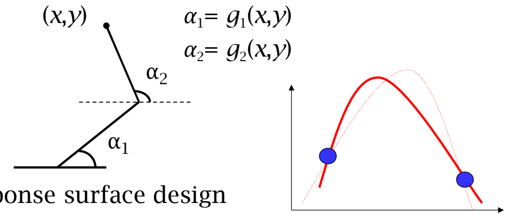
- Example: Price of a used car
- x : car attributes
  y : price
  $y = g(x \mid \vartheta)$
  $g(\ )$ model,
  ϑ parameters



33

Regression Applications

- Navigating a car: Angle of the steering wheel (CMU NavLab)
- Kinematics of a robot arm



Response surface design

34

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

35

Unsupervised Learning

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

36

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, ...

37

Resources: Datasets

- UCI Repository: http://archive.ics.uci.edu/ml/
- MNIST Database: ttp://yann.lecun.com/exdb/mnist/
- UCI KDD Archive: http://kdd.ics.uci.edu/summary.data.application.html
- A list ... http://www.dmoz.org/Computers/Artificial_Intelligence/Machine_Learning/Datasets/
  - Statlib: http://lib.stat.cmu.edu/
  - Delve: http://www.cs.utoronto.ca/~delve/

38

Resources: Journals

- Journal of Machine Learning Research www.jmlr.org
- Machine Learning
- Neural Computation
- Neural Networks
- IEEE Transactions on Neural Networks
- IEEE Transactions on Pattern Analysis and Machine Intelligence
- Annals of Statistics
- Journal of the American Statistical Association
- ...

39

Resources: Conferences

- International Conference on Machine Learning (ICML)
  - ICML05: http://icml.ais.fraunhofer.de/
- European Conference on Machine Learning (ECML)
  - ECML05: http://ecmlpkdd05.liacc.up.pt/
- Neural Information Processing Systems (NIPS)
  - NIPS05: http://nips.cc/
- Uncertainty in Artificial Intelligence (UAI)
  - UAI05: http://www.cs.toronto.edu/uai2005/
- Computational Learning Theory (COLT)
  - COLT05: http://learningtheory.org/colt2005/
- International Joint Conference on Artificial Intelligence (IJCAI)
  - IJCAI05: http://ijcai05.csd.abdn.ac.uk/
- International Conference on Neural Networks (Europe)
  - ICANN05: http://www.ibspan.waw.pl/ICANN-2005/
- ...

40

Schedule

HW1 –SVM	• Linear Classifiers – KNN and SVM	
	• Decision Trees	Project Selection
HW2– RDF	• Ensemble Learning	
HW3 – PCA	• Feature Selection and Reduction	
HW4 – NN	• Neural Networks	
	• Bayesian Learning	Midterm
HW5 – Clustering	• Clustering	
	• Manifold Learning	
HW5 – RL	• Reinforcement Learning	Project Presentation
	• Learning Theory	Final

41

Linear Classifiers

- Most of SVM related slides are adapted from Mingyue Tan of the University of British Columbia

42

Linear Classifiers – Notation

Given: Observations $x_i \in R^n, i = 1, \dots, l$ and associated truths y_i by a trusted source

Goal: Construct a machine/algorithm that learns the mapping $\boldsymbol{x}_i \to y_i$

The deterministic machine is defined by a set of possible mapping $\boldsymbol{x}_i \to f(\boldsymbol{x}, \alpha)$.

A particular choice of α generates a "trained machine".

A Tutorial on SVM for PR by Burges, 1998

43

Supervised Machine Learning

of experiences | observed outcome | observed input

$$\arg\max_{f \in F} \sum_{i=1}^{n} goodness(y_i, f(x_i))$$

hypothesis/program space | performance measure | selected hypothesis

44

Supervised Machine Learning

of experiences · observed outcome · observed input

$$\arg\max_{\alpha} \sum_{i=1}^{n} goodness(y_i, f(\alpha, x_i))$$

hypothesis/program space · performance measure · selected hypothesis

45

Notation

Example: Tree recognition problem…

- Observations $\boldsymbol{x}_i \in R^{256}, i = 1, \dots, l$ comprised of 16x16 image windows representing trees.
- Associated truths $y_i \in \{1, -1\}$ for where 1 is for tree and -1 is no-tree images.
- This is **two-class pattern-recognition** (or classification)…

A Tutorial on SVM for PR by Burges, 1998

46

Supervised Machine Learning

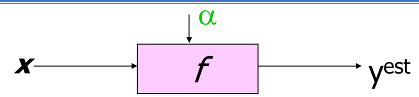
of experiences · observed outcome · observed input

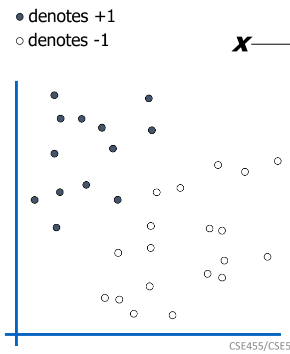
$$\arg\min_{\alpha} \sum_{i=1}^{n} diff(y_i, f(\alpha, x_i))$$

hypothesis/program space · performance measure · selected hypothesis

47

Simple Classification Problem



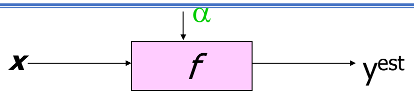


How would you classify this data?

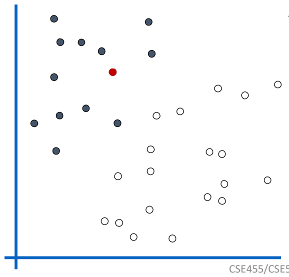
48

Simple Classification Problem

- denotes +1
- denotes -1



$f(\boldsymbol{x}, \{\boldsymbol{x}_1,\dots,\boldsymbol{x}_n\}) = \boldsymbol{x}_i$ in $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_n\}$ closest to $\boldsymbol{x}$.

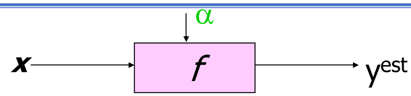


Naïve classifier: look at the closest point…

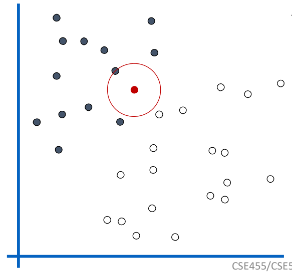
49

Nearest Neighbor

- denotes +1
- denotes -1



$f(\boldsymbol{x}, \{\boldsymbol{x}_1,\dots,\boldsymbol{x}_n\}) = \boldsymbol{x}_i$ in $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_n\}$ closest to $\boldsymbol{x}$.

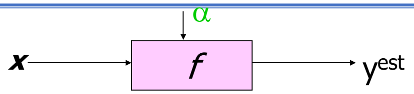


Questions:
- What's closest? $d_i(\boldsymbol{x}) = |\boldsymbol{x} - \boldsymbol{x}_i|_2$
- How to implement?
  - Smooth?
- Tractable?

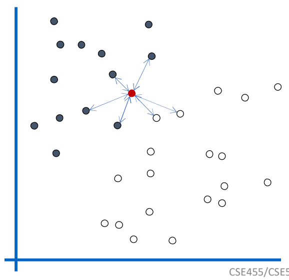
50

KNN

- denotes +1
- denotes -1



$f(\boldsymbol{x}, k, \{\boldsymbol{x}_1,\dots,\boldsymbol{x}_n\}) = \frac{1}{k}\sum_{x_i \in N_k(x)} y_i$



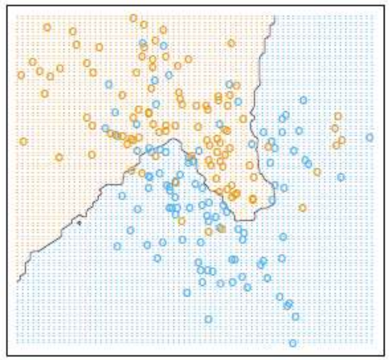
k-nearest neighbor fit…

$N_k(x)$ is the neighborhood of $\boldsymbol{x}$ defined by the k closest points $\boldsymbol{x}_i$ in the training set

Voronoi diagrams…

51

KNN

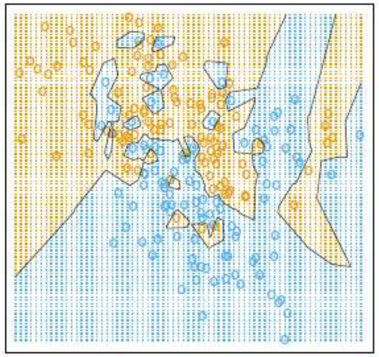


k=15

Hastie et al.

52

KNN vs NN



k=1

Hastie et al.

53

Implementation

- 1-nearest neighbor
  - Applied to large % of low-dimensional problems
- Kernel methods
  - Use weights decreasing smoothly to zero with distance from the target point, rather than the effective 1/0 weights used by k-nearest neighbors
- High dimensions
  - Use distance metrics emphasizing some variables more than others

54

Thanks for listening!

55