Intro to AI: A Statistical Approach (aka Machine Learning)

Instructor: Taylor Berg-Kirkpatrick
Slides: Sanjoy Dasgupta
Course website: http://cseweb.ucsd.edu/classes/fa19/cse151-a/

Machine learning versus Algorithms

A central goal of both fields:

develop procedures that exhibit a desired input-output behavior.

- **Algorithms**: input-output mapping can be precisely defined.
  Input: Graph G, two nodes u, v in the graph.
  Output: Shortest path from u to v in G

- **Machine learning**: mapping cannot easily be made precise.
  Input: Picture of an animal.
  Output: Name of the animal.

Machine learning versus Algorithms

A central goal of both fields:

develop procedures that exhibit a desired input-output behavior.

- **Algorithms**: input-output mapping can be precisely defined.
  Input: Graph G, two nodes u, v in the graph.
  Output: Shortest path from u to v in G

- **Machine learning**: mapping cannot easily be made precise.
  Input: Picture of an animal.
  Output: Name of the animal.

Instead, provide examples of (input,output) pairs.
Ask the machine to *learn* a suitable mapping itself.

Prediction problems: inputs and outputs

Basic terminology:

- The input space, $\mathcal{X}$.
  E.g. 32×32 RGB images of animals.
- The output space, $\mathcal{Y}$.
  E.g. Names of 100 animals.

x:



y: "bear"

Prediction problems: inputs and outputs

Basic terminology:

- The input space, $\mathcal{X}$.
  E.g. 32×32 RGB images of animals.
- The output space, $\mathcal{Y}$.
  E.g. Names of 100 animals.

x:



y: "bear"

After seeing a bunch of examples (x, y), pick a mapping

$$f : \mathcal{X} \to \mathcal{Y}$$

that accurately recovers the input-output pattern of the examples.

Prediction problems: inputs and outputs

Basic terminology:

- The input space, $\mathcal{X}$.
  E.g. 32×32 RGB images of animals.
- The output space, $\mathcal{Y}$.
  E.g. Names of 100 animals.

x:



y: "bear"

After seeing a bunch of examples (x, y), pick a mapping

$$f : \mathcal{X} \to \mathcal{Y}$$

that accurately recovers the input-output pattern of the examples.

Categorize prediction problems by the type of **output space**:
(1) discrete, (2) continuous, or (3) probability values

Discrete output space: classification

Binary classification

E.g., Spam detection
$\mathcal{X} = \{\text{email messages}\}$
$\mathcal{Y} = \{\text{spam}, \text{not spam}\}$

Discrete output space: classification

Binary classification

E.g., Spam detection
$\mathcal{X} = \{\text{email messages}\}$
$\mathcal{Y} = \{\text{spam}, \text{not spam}\}$

Multiclass

E.g., News article classification
$\mathcal{X} = \{\text{news articles}\}$
$\mathcal{Y} = \{\text{politics}, \text{business}, \text{sports}, \ldots\}$

Discrete output space: classification

Binary classification

E.g., Spam detection
$\mathcal{X} = \{\text{email messages}\}$
$\mathcal{Y} = \{\text{spam}, \text{not spam}\}$

Multiclass

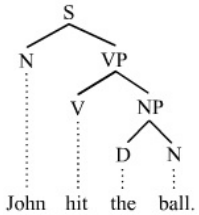
E.g., News article classification
$\mathcal{X} = \{\text{news articles}\}$
$\mathcal{Y} = \{\text{politics}, \text{business}, \text{sports}, \ldots\}$

Structured outputs

E.g., Parsing
$\mathcal{X} = \{\text{sentences}\}$
$\mathcal{Y} = \{\text{parse trees}\}$

$x =$ "John hit the ball"

$y =$



Continuous output space: regression

- Pollution level prediction
  Predict tomorrow's air quality index in my neighborhood
  $\mathcal{Y} = [0, \infty)$ (< 100: okay, > 200: dangerous)

- Insurance company calculations
  What is the expected life expectancy of this person?
  $\mathcal{Y} = [0, 120]$

What are suitable predictor variables ($\mathcal{X}$) in each case?

Probability estimation

$\mathcal{Y} = [0, 1]$ represents **probabilities**

Example: Credit card transactions

- $x =$ details of a transaction
- $y =$ probability this transaction is fraudulent

Probability estimation

$\mathcal{Y} = [0, 1]$ represents **probabilities**

Example: Credit card transactions

- $x =$ details of a transaction
- $y =$ probability this transaction is fraudulent

Why not just treat this as a binary classification problem?

Perequisites

- Probability
  Conditional probability, Bayes' rule, random variables, independence

Perequisites

- Probability
  Conditional probability, Bayes' rule, random variables, independence

- Linear algebra
  Matrix-matrix product, rank, matrix inverse, determinant, eigen decomp

Perequisites

- Probability
  Conditional probability, Bayes' rule, random variables, independence
- Linear algebra
  Matrix-matrix product, rank, matrix inverse, determinant, eigen decomp
- Vector calculus
  Multivariate differentiation, partial gradients, chain rule, Hessian

Perequisites

- Probability
  Conditional probability, Bayes' rule, random variables, independence
- Linear algebra
  Matrix-matrix product, rank, matrix inverse, determinant, eigen decomp
- Vector calculus
  Multivariate differentiation, partial gradients, chain rule, Hessian
- Coding experience
  Python, Numpy, PyTorch

Roadmap for the course

1. Introduction / Prediction problems
   Classification, regression, probability estimation

Roadmap for the course

1. Introduction / Prediction problems
   Classification, regression, probability estimation

2. Generative models for classification
   Maximum-likelihood estimation, Gaussian generative models

Roadmap for the course

1. Introduction / Prediction problems
   Classification, regression, probability estimation

2. Generative models for classification
   Maximum-likelihood estimation, Gaussian generative models

3. Linear prediction
   Linear regression, logistic regression, perceptron, support vector machines

Roadmap for the course

1. Introduction / Prediction problems
   Classification, regression, probability estimation

2. Generative models for classification
   Maximum-likelihood estimation, Gaussian generative models

3. Linear prediction
   Linear regression, logistic regression, perceptron, support vector machines

4. Non-linear models / Deep learning
   Nearest neighbors, decision trees, neural networks

Roadmap for the course

1. Introduction / Prediction problems
   Classification, regression, probability estimation

2. Generative models for classification
   Maximum-likelihood estimation, Gaussian generative models

3. Linear prediction
   Linear regression, logistic regression, perceptron, support vector machines

4. Non-linear models / Deep learning
   Nearest neighbors, decision trees, neural networks

5. Representation learning
   Clustering, PCA, embeddings, autoencoders

Quizzes / Homeworks

- Quizzes
  One quiz for each of the five sections
  Announced in advance and taken in class, drop lowest

Quizzes / Homeworks

- Quizzes
  One quiz for each of the five sections
  Announced in advance and taken in class, drop lowest

- Homeworks
  One HW for each of the five sections + warmup HW
  Mixture of math and coding problems using PyTorch
  Use latex and turn in pdf Graded on random subset of problems, drop lowest

Quizzes / Homeworks

- Quizzes
  One quiz for each of the five sections
  Announced in advance and taken in class, drop lowest

- Homeworks
  One HW for each of the five sections + warmup HW
  Mixture of math and coding problems using PyTorch
  Use latex and turn in pdf Graded on random subset of problems, drop lowest

- Grading
  50% Quizzes / 50% HWs. No midterm! No final!

Quizzes / Homeworks

- Quizzes
  One quiz for each of the five sections
  Announced in advance and taken in class, drop lowest

- Homeworks
  One HW for each of the five sections + warmup HW
  Mixture of math and coding problems using PyTorch
  Use latex and turn in pdf Graded on random subset of problems, drop lowest

- Grading
  50% Quizzes / 50% HWs. No midterm! No final!

- Getting help
  Come to office hours and discussion!
  We will help your learn PyTorch!
  Join Piazza!