Machine Learning

Lecture 28: Closing

Hao Tang

November 21, 2022

Critical thinking

- What are the assumptions?
  - What are the properties?
  - Survey?
- Can I make up assuptions?
  - What conclusions do I want?
  - Can I work backwards?
  - What is the ideal case?
  - What is the worst case?
- Where are the assumptions used?

Types of thinking

- Mathematical
  - What happens if we make these assumptions?
- Computational
  - What are the exact steps?
- Statistical
  - How many samples do we need?

Critical thinking

- Compare discriminative and generative approaches.

Critical thinking

- Compare discriminative and generative approaches.
- Compare logistic regression, perceptron, and support vector machines.

Critical thinking

- Compare discriminative and generative approaches.
- Compare logistic regression, perceptron, and support vector machines.
- Compare logistic regression and 6-layer neural network with 1024 units.

Machine learning

- What is machine learning?

Machine learning

- What is machine learning?
- When to use machine learning?

Machine learning

- What is machine learning?
- When to use machine learning?
- How to use machine learning?

Connections to other courses

- Foundations of Data Science (FDS)
- Applied Machine Learning (AML)
- Machine Learning and Pattern Recognition (MLPR)
- Probabilistic Modeling and Reasoning (PMR)
- Machine Learning Practical (MLP)
- Machine Learning Theory (MLT)
- Reinforcement learning (RL)

Connections to other courses

- Foundations of natural language processing (FNLP)
- Accelerated natural language processing (ANLP)
- Natural language understanding, generation, and machine translation (NLU+)
- Speech processing (in PPLS)
- Automatic speech recognition (ASR)
- Speech synthesis (in PPLS)
- Image and vision computing
- Advanced robotics

Tasks we haven't talked about

- Information retrieval
- Recommendation system
- 3D reconstruction
- Text generation
- Protein folding

Evaluation we haven't talked about

- Mean average precision
- Receiver operating characteristic (ROC) curve
- Word error rates (WER)
- Bilingual evaluation understudy (BLEU) score
- Mean opinion score (MOS)

Techniques we haven't talked about

- More optimization
  - There are a lot of problems that cannot be solved with gradient descent.

Techniques we haven't talked about

- More optimization
  - There are a lot of problems that cannot be solved with gradient descent.
- k nearest neighbor
  - A point is like its neighbors.

Techniques we haven't talked about

- More optimization
  - There are a lot of problems that cannot be solved with gradient descent.
- k nearest neighbor
  - A point is like its neighbors.
- Decision trees
  - Decisions are based on conditional statements.

Techniques we haven't talked about

- More optimization
  - There are a lot of problems that cannot be solved with gradient descent.
- k nearest neighbor
  - A point is like its neighbors.
- Decision trees
  - Decisions are based on conditional statements.
- Dimensionality reduction
  - Points in space have structures.

Techniques we haven't talked about

- More optimization
  - There are a lot of problems that cannot be solved with gradient descent.
- k nearest neighbor
  - A point is like its neighbors.
- Decision trees
  - Decisions are based on conditional statements.
- Dimensionality reduction
  - Points in space have structures.
- Bayesian approaches
  - Beliefs are updated and marginalized.

Learning settings we haven't talked about

- Active learning
  - What is the next sample that is most useful to learn?

Learning settings we haven't talked about

- Active learning
  - What is the next sample that is most useful to learn?
- Online learning
  - What can we learn if we make decisions sequentially and can look back?

Learning settings we haven't talked about

- Active learning
  - What is the next sample that is most useful to learn?
- Online learning
  - What can we learn if we make decisions sequentially and can look back?
- Semi-supervised learning
  - What do we do with unlabeled data?

Learning settings we haven't talked about

- Active learning
  - What is the next sample that is most useful to learn?
- Online learning
  - What can we learn if we make decisions sequentially and can look back?
- Semi-supervised learning
  - What do we do with unlabeled data?
- Federated learning
  - Can learning be distributed?

Aspects

- Privacy
- Fairness
- Interpretability

Differentially private

A learning algorithm A is differentially private if for all data sets S and $S^{(i)}$ that differs in the i-th sample,

$$\mathbb{P}[A(S) \in \mathcal{G}] < e^{\epsilon}\mathbb{P}[A(S^{(i)}) \in \mathcal{G}] + \delta \tag{1}$$

for any subset $\mathcal{G} \subseteq \mathcal{H}$ of hypothesis.

Fairness

- An outcome Y should be independent of groups G. ($Y \perp G$)
- A score R should be independent of the groups G given the outcome Y. ($R \perp G | Y$)
- An outcome Y should be independent of the groups G given the score R. ($Y \perp G | R$)

Artificial Intelligence

- Building models of intelligence, not solving tasks
- Involving philosophy, psychology, linguistics, etc

What we have learned in this course

- Formalize a ML problem with math
- Read and understand ML theorems
- Turn ML algorithms into programs