CS 5806 Machine Learning II

Lecture 3 - Learning Strategies
August 28th, 2023
Hoda Eldardiry

Reminders

- Join the intro discussion
- Signup on the project teaming googlesheet
- Form your team
- Reinforcement Learning (Lecture 1 cont.)

Lecture Objectives

- Theoretical learning strategies

- Practical learning strategies

Theoretical Learning Strategies

Learning theory

- In **what cases** can we learn?
- **How well** can we learn?
- **How many examples** do we need to learn?
- How do we **quantify** our ability to **generalize** to unseen data?
- **Which algorithms** are better suited to specific learning **settings**?

Hypothesis Space

- **<u>Hypothesis space *H*</u>**
  - Set of all hypotheses h that the learner may consider
  - **<u>L*earning* *is a search*</u> *through hypothesis space***

- **Objective**
  - Find h that minimizes some error function w.r.t. training examples
  - Error function example: misclassification

- What about unseen examples?

Generalization

- Recall
  - A **good** hypothesis will **generalize well** (predict unseen examples correctly)
  - Prefer the simplest hypothesis consistent with the data

- Typically
  - If hypothesis h approximates target function ***f*** well over a sufficiently large training set
  - Then h will also approximate **f** well over unobserved examples

- Therefore
  - **A sufficient training set size enables generalization**

- <u>Why do you think that is the case?</u>

Training Set Size & Generalization

- Why does a sufficient training set size enables generalization?

Inductive Learning

- **Goal**
  Find an h that agrees with ***f*** on training set
  h is **consistent** if it agrees with ***f*** on all examples

- Finding a consistent hypothesis is not always possible

  - **Insufficient hypothesis space**
    E.g., it is not possible to learn exactly f(x) = ax+b+xsin(x)
    when ***H*** = space of polynomials of finite degree

  - **Noisy data**
    E.g., in weather prediction:
    identical conditions may lead to rainy and sunny days

Inductive Learning

- A learning problem is **realizable** if the hypothesis space contains the true function — otherwise it is **unrealizable**

- It is difficult to determine whether a learning problem is realizable since the true function is not known

- It is possible to use a very large hypothesis space
  – For example: H = class of all Turing machines

- But there is a **tradeoff** between **expressiveness** of a hypothesis class and the **complexity** of finding a good hypothesis

Practical learning strategies

WHERE DO WE START?

Where do we start?

- We identify a problem
- We want to use Machine Learning to solve it
- Where do we start?
- What questions would you ask?

What is possible? (ML expertise)

- **Model**
  - Problem solved (published)?
  - A model for this problem?
  - Improve an existing model?
  - A different approach?
  - Model complexity?
  - Model implementation/deployment challenges?
  - Is ML necessary/useful?

What is possible? (ML expertise)

- **Data**
  - What data can our model use?
    - Balanced/unbalanced?
    - Size?
    - Available/accessible?
    - Characteristics (modality, multi-source, real-time, dynamic)?
  - How can we preprocess the data?
    - Filter/clean/label/remove noise?
    - Resources available to enable learning at such a data size?
    - Features? Feature Engineering?

What is needed?
(Domain expertise)

- Problem?
- Goals? Predict, Classify, Automate?
- Envisioned end result?
- Schedule and budget for deliverables?
- Objectives?
- Constraints?
- Trade-offs?

What is needed?
(Domain expertise)

- **Data**
  - What available domain knowledge can be leveraged?
  - Could data be simulated?
  - What data can we collect?
  - Is it free or do we need to pay for it?

What are the trade-offs? (Decisions & Choices)

- How much accuracy is needed?
- How do we evaluate success?
- What are the computing resources needed/available?
- Will we allow the solution...
  - Unlimited runtime?
  - Unlimited space?

ML & ETHICS

Responsible AI-ers

- From ethical principles to policies

- Formulating policies into algorithmic specifics

From ethical principles to policies

From ethical principles to policies

- What **ethical issues** are associated with the development and deployment of AI systems?

- What **ethical principles** can be applied to address these issues?

Ethical Considerations for AI Design

- **Impact**
  - Social & ethical implications
  - Impact on users & society
  - Impact on daily human life

- **Benefits & Risks**
  - Will both be distributed equitably?
  - What can an AI designer do about benefits & risks?
  - E.g., how much access to data can an AI algorithm have?

- **Autonomous Decision Making**
  - Are we designing autonomous decision making?
  - Are we enabling autonomously making decisions that affect others
  - How much of a decision can humans rely on AI to make?

Examples issues associated with AI development & deployment

- Decision-making design
- Restricting intelligent agents that interact with the world
- Privacy
- Output augmentation for transparency, trust, & collaborative learning

Decision-Making Design

- Social & ethical implications of AI models making decisions that affect people?

- Example
  — Self-driving cars use deep neural networks to make decisions & rely on sensors
  — Sensor provide perception of surroundings
  — Learn through experience
  — Not driven by possible outcomes
  — How decisions made when all possible outcomes are negative?
  — A policy should guide what to do next
  — Design it in a way that the vehicle immediately stops to avoid any harm?

- Decision should be based on least negative outcome

- How can AI designer pick the better option? Based on what criteria?

- Implications of chosen method?

Restrict intelligent agents that interact with the world

- Interactions between intelligent agents & humans must be controlled for human safety
- Incorporate ethical boundaries that AI should not cross
  - As AI gets more sophisticated
  - As AI approaches true general intelligence
  - As errors can cause harm or break social laws

Privacy

- AI solutions should preserve privacy of sampled individuals
- Any data that can reveal identity should be protected
- Examples:
  — A single attribute (e.g., school ID number)
  — Attribute pairs (e.g.,“location at 2 am”, “location at 2 pm”)
  — A virtual/augmented reality (VR/AR) hand-tracking system can violate privacy if combined with biometric data
- Privacy is not guaranteed when certain combinations of data reveals identity
- Individuals should be informed about:
  — How the data might become public
  — What the risks are
  — How likely it is that such risks will happen

Data Privacy

- Attackers can manipulate data so AI algorithm produce biased results

- AI algorithms should guarantee data privacy

- To address data compromise
  - Detect anomalies
  - Identify irregular results

Output augmentation for transparency, trust & collaborative learning

- Primary AI algorithm output:
  — Decision recommendations, predictions

- Additional output:
  — Explanations, variable importance, assumptions

- Additional output provides transparency so user:
  — Can trust the output
  — Can evaluate the output based on:
    — Context-awareness
    — Domain expertise
    — Ethical considerations

Issues to be explicitly addressed

- Incentive
- Professional responsibility

Incentive

- AI designers should understand:
  — Context of system
  — Consequences

- Example (monetary incentives):
  — Search engine trained to generate news search results that optimize for **ad revenue**
  — Can lead to gender/racial/socioeconomic bias

  — If **money** is the driving factor in what & how news is reported
  — Clickbait titles & outrage-inducing stories could become the norm
  — Potentially hindering healthy civil debates

Professional responsibility

- In general: AI professionals have social & ethical responsibility to
  - Contribute to society
  - Avoid harm
  - Be honest & trustworthy
  - Be fair
  - Respect privacy

- In particular: AI designers should
  - Interpret results without judgment or bias
  - Use data samples representative of the full population
  - Build models that do not
    - Break the law
    - Exacerbate existing problems
    - Introduce bias
    - Cause discrimination against protected classes

Formulating policies into algorithmic specifics

Formulating policies into algorithmic specifics

- How to develop AI systems in an environment with various constraints?

- Ethical AI design considerations & trade-offs
  - Generalization versus specificity
  - Optimizing decision versus competition harm
  - Personalization versus privacy/bias
  - Safety versus automation

Formulating policies into algorithmic specifics

Generalization versus specificity

- Should AI solutions be general in order to be less biased?

- What if generalization reduces accuracy?

- Example
  — A model that detects a highly-lethal disease in people
  — If certain genetics cause the disease to interact with the body differently
  — Our ML model must learn those differences & not lose them for the sake of generality

Formulating policies into algorithmic specifics

Optimizing decision versus competition harm

- Example: Clustering & classifying stock market trends to explore patterns & relationships
  — Leveraging AI insights in a trading strategy
  — Ethical consideration: leveraging the data to cause harm to other investors
  — Governing policy: put limits on sell orders, to not crash a stock price

- However, it's rare that an independent agent can drastically affect the price of an asset

- Counterpoint to consider: using AI technology on the stock market as a whole

- Example: if enough people are using similar AI models to trade
  — Then market prices are driven more by AI agents speculating on stochastic processes
  — Rather than on human investor judgment

- Example: the market can have ``flash crashes" due to algorithms selling off stock in huge quantities because other algorithms are selling as well
  — Ultimately, the stock market prices affect a lot of people's retirement funds, so if an AI agent damages the market severely, it would have widespread implications

What about your project topic?

- What ethical aspects may arise?

- What can you do about them?