Ling/CSE 472: Introduction to Computational Linguistics

April 14
ML

Overview

- foma examples
- ML
  - Big picture / then zooming out
  - A few key points
- Reading questions

Machine learning, in a nutshell

- "Once the three components ⟨T,P,E⟩ have been specified fully, the learning problem is well defined" (Mitchell, p.2)

Task definition	Learning approach
Evaluation metric	Train/test data

Machine learning, in context

Why do we care about this task?

How does dataset model the task?

-build something useful
-learn about: computers, people, modeling domain

Task | Lear
Eval | Train

How does the task relate to the world?

How do we collect the data?

What happens when we deploy this?

What is ML (Tom Mitchell)

- Study of algorithms that:
    - improve their performance
    - at some task
    - with experience
- Data → ML → **Understanding?**

What is learning? (one definition)

- A process in which:
  - You observe
  - You extrapolate some knowledge from the observations
  - ...so that you can predict something about the future observations
- What is **machine** learning?

What is machine learning?

- It is a process in which:
  - You represent observations as specific (x,y) pairs (x can be a vector of values)
  - You extrapolate some *mathematical function* from the observations
  - ...You can now feed to the function new data points, mapped to the same x representations, and it outputs y
- But why? A mathematical equation is not real, is it? Why do we use it to represent the world? Why vectors, too?

ML Tasks: Classification

- From data to discrete labels
    - Spam filtering
    - Text classification
    - Object detection
    - Weather prediction (e.g. rain, snow...)
    - Sentiment analysis
    - etc.

ML Tasks: Regression

- Predict a numeric value
    - Stock market
    - Weather prediction (temperature)
    - Airfares

ML Tasks: Similarity

- Finding data
    - Given image, find similar ones
    - Similar products, songs...
    - Similar texts
    - Similar words...

Machine Learning: Evaluation

- Human evaluation: How good is this output?
  - E.g. scores of fluency/adequacy for MT output
- Automatic evaluation: How well does this output match the stored gold standard?
  - Test data should be ‘held out’ (considered by neither the algorithm nor the developer)
    - Tests trained model’s ability to ‘generalize’
  - Metrics must be computable by comparing machine output to ground truth labels

Train/dev/test splits

- “Test” is held out: don’t look, just report numbers
    - The parable of Sec 23 of the WSJ
- Train: training data, for training up the model
- Dev: development data
    - Parameter tuning
    - Error analysis
- Really small datasets: cross-validation

Mitchell’s (2017) key questions in ML

- How can computers improve performance through experience?
- Which theoretical laws govern learning systems?
- What are the key questions in NLP? Computational linguistics? Linguistics?

ML: Key results

- No free lunch
    - “...no system has any basis to reliably classify new examples that go beyond those it has already seen...”
- Three sources of error: Bias, variance, and unavoidable error:
    - some probability of us being wrong
- Overfitting
    - When true error > train error
    - What is the relationship between true error and test error?

Where is ML headed next?

- Will ML change the way we think about human learning?
- Human-machine (learning) interaction
- ML by reading
- Note that both directions involve natural language understanding

Reading questions

- In the key concepts section, probability is mentioned a lot when discussing the training algorithms. What does probability have to do with training a machine? Is it for selecting random data? Or something else?

- What is the general meaning of the terms "generative" and "discriminative"?

- In the context of machine learning, what is a hypothesis? Is a hypothesis simply another term for an algorithm produced by machine learning? What does the set of all possible hypotheses represent?

Reading questions

- Unavoidable error is in the name of the error, but it seems to be there because the machine is making a validity decision based on probability. Is it theoretically possible to raise the certainty to the point there is no "unavoidable error"?

- I have a question about the section on overfitting: how is a hypothesis' true error calculated? Is true error just the error rate that occurs when given any collection of data (not just the training data)?

- If regularization increases bias in order to reduce learning algorithm's sensitivity to variance in training examples, isn't this capable of causing more errors even though it is meant to be a solution to overfitting? Is there a real-life example of how regularizations are used?

Reading questions

- I found the discussion of reinforcement learning to be the most interesting, how does one show a computer that something is positive or negative. What does it consider a "reward"?

- What does the perspective Machine Learning as evolutionary search mean? That machines learning evolves and improves based on basic evolutionary principles like survival of the fittest?

- How is it that learning many inter-related functions can be easier than a single target? Is it that there's some additional context or information that's provided in this way? And if that's the case, and we know that learning many functions at once can be easier, then what's preventing us from shifting to training machine learning agents that learn multiple functions at once?

Reading questions

- I'm a bit confused at how learning is defined in the paper. Is it defined in this case as just a general algorithm that improves automatically, and, if so, what is the biggest goal in NLP currently? Just improvement of machines' interpretation of human language?

Reading questions

- How has field-specific schema survived with modern machine learning, specifically with linguistics? One common sentiment I hear from researchers in NLP with a formal linguistic background is that the lack of linguistic consideration (outside of BERT/mBert...) has been disappointing. It is clear that reflecting on the prior schema in a field has potential, but the "uninterpretable" status quo has infiltrated most ML research.

Reading questions

- What kind of specific tasks can a Neural Net can solve in NLP? The segmentation problem in ASR? TTS using wave-net? Can it be also used for some sort of semantic processing components? (I know this might be a stretch, but can it "understand" the meaning, or at least mimic the behavior of an entity that understands a language?)

- One of the things I'm most curious about is what range of tasks can a neural network solve in NLP? Is linguistic consideration expandable, if this range is limited?

Reading questions

- If machines were to learn by reading large portions or even the entirety of the internet, how would potentially false information be discovered and dealt with?