Lecture 1, COMS E6998, Spring 2012

Michael Collins

January 23, 2012

Today's Lecture

- Introduction:
  - Example problems from machine learning for NLP
  - Topics we'll cover in the course
  - Background required for the course
  - Projects/homework assignments
- Hidden Markov models
- Discriminative models for classification

A Machine-Learning Example: Hand-Written Digit Recognition

- The problem: given a hand-written digit, decide whether it is $0, 1, 2, \ldots$ or 9
- A learning approach:
  1. Collect several hundred/thousand example digits, and label them by hand to form a *training set*
  2. Automatically learn a digit recognition *model* from the training set
  3. Apply the model to new, previously unseen hand-written digits
- Systems built in this way are in widespread use in the U.S. postal service (ZIP-code recognition), and in automatic check-reading

Related Problems

- Identifying faces within an image
  (see the Viola and Jones face detector)
- Text classification/spam filtering
- Medical applications: e.g., classification of cancer type
- Information retrieval: e.g., ranking web-pages in order of relevance to a given query

Supervised Learning Problems

- Goal: Learn a function $f : \mathcal{X} \rightarrow \mathcal{Y}$

- We have n training examples

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$

  where each $x_i \in \mathcal{X}$, and each $y_i \in \mathcal{Y}$

- Often (not always) $\mathcal{X} = \mathbb{R}^d$ for some integer d

- Some possibilities for $\mathcal{Y}$:
  - $\mathcal{Y} = \{-1, +1\}$ (binary classification)
  - $\mathcal{Y} = \{1, 2, \ldots, k\}$ for some $k > 2$ (multi-class classification)
  - $\mathcal{Y} = \mathbb{R}$ (regression)

Sequence Labeling Problems

- Task: learn a function that maps an input sequence

  $$x_1, x_2, \ldots, x_m$$

  to an output sequence

  $$y_1, y_2, \ldots, y_m$$

  Note: each $y_i \in \mathcal{Y}_i$ where $\mathcal{Y}_i$ is a **finite** set of possible labels at the i'th position

- This is a core problem in natural language processing

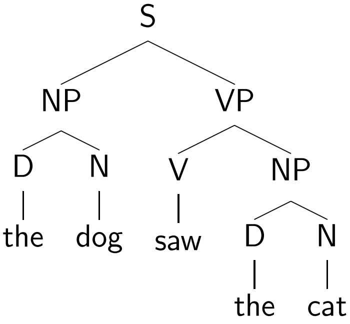
- Examples: part-of-speech tagging, named-entity recognition

Context-Free Parsing

- The task: learn a function that maps a sentence, e.g.,

  the dog saw the cat

  to a parse tree,

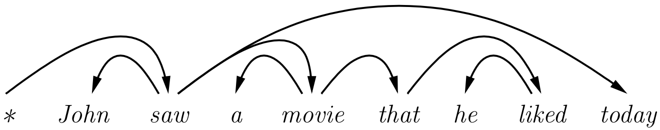


Dependency Parsing

- The task: learn a function that maps a sentence, e.g.,

  John saw a movie that he liked today

  to a *dependency structure*,



Machine Translation

- The task: learn a function that maps a sentence in one language, e.g.,

  In wenigen Tagen finden Parlamentswahlen in Slowenian statt

  to a sentence in another language,

  In a few days elections take place in Slovenia

Mapping Sentences to Logical Form

- The task: learn a function that maps a sentence e.g.,

  Show me the latest flight from Boston to Seattle on Friday

  to a expression in logical form that represents its meaning, e.g.,

  $argmax(\lambda x.flight(x) \wedge from(x, BOS) \wedge to(x, SEA) \wedge day(x, FRI), \lambda y.time(y))$

Topics Covered in the Class

- Probabilistic models for structured NLP data
  - e.g., hidden Markov models (HMMs), maximum-entropy Markov models (MEMMs), conditional random fields (CRFs), probabilistic context-free grammars, synchronous context-free grammars, dependency parsing models, etc.
- Semi-supervised learning
  - e.g., the EM algorithm, deriving lexical representations from unlabeled data, cotraining, entropy regularization, canonical correlation analysis (CCA)
- Inference algorithms
  - e.g., dynamic programming, belief propagation, methods based on linear programming and integer linear programming, dual decomposition/Lagrangian relaxation

Admin

- Background required for the class: a prior class in machine learning and/or natural language processing
- Strong background in algorithms, and probability/statistics

- Evaluation:
  - Final class project (40%)
  - 3 homeworks (30%)
  - One 2 hour exam (in class, date TBD) (30%)