CS778 Seminar
Learning to Predict Structured Objects

Fall 2006

Thorsten Joachims

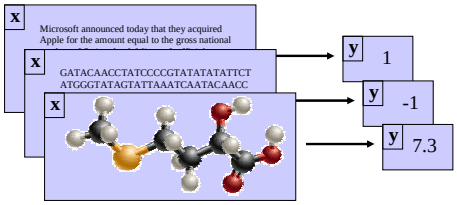
Cornell University
Department of Computer Science

Supervised Learning

- **Find function from input space *X* to output space *Y***

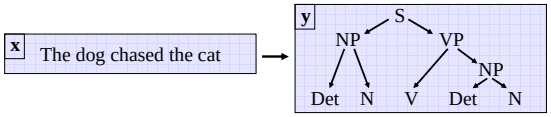
$$h : X \longrightarrow Y$$

such that the prediction error is low.



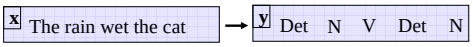
Examples of Complex Output Spaces

- **Natural Language Parsing**
  - Given a sequence of words *x*, predict the parse tree *y*.
  - Dependencies from structural constraints, since *y* has to be a tree.



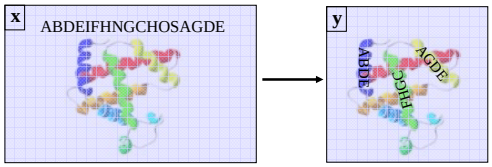
Examples of Complex Output Spaces

- **Part-of-Speech Tagging**
  - Given a sequence of words *x*, predict sequence of tags *y*.
  - Dependencies from tag-tag transitions in Markov model.



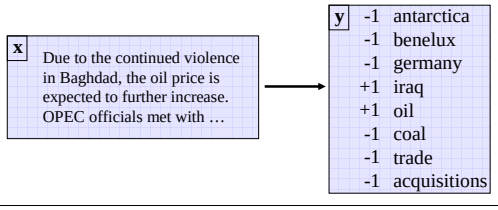
Examples of Complex Output Spaces

- **Sequence to Structure Alignment**
  - Given a sequence/structure pair x=(x_1,x_2), predict the correct sequence of alignment operations *y*.
  - Dependencies because operations are applied sequentially



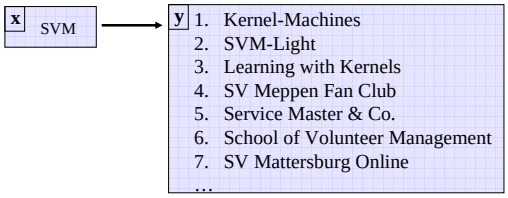
Examples of Complex Output Spaces

- **Multi-Label Classification**
  - Given a (bag-of-words) document x, predict a set of labels *y*.
  - Dependencies between labels from correlations between labels ("iraq" and "oil" in newswire corpus)



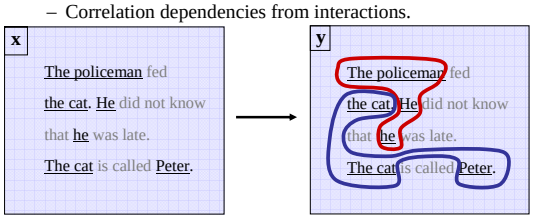
Examples of Complex Output Spaces

- **Information Retrieval**
  - Given a query x, predict a ranking *y*.
  - Dependencies between results (e.g. avoid redundant hits)
  - Loss function over rankings (e.g. AvgPrec)



Examples of Complex Output Spaces

- **Noun-Phrase Co-reference**
  - Given a set of noun phrases *x*, predict a clustering *y*.
  - Structural dependencies, since prediction has to be an equivalence relation.
  - Correlation dependencies from interactions.

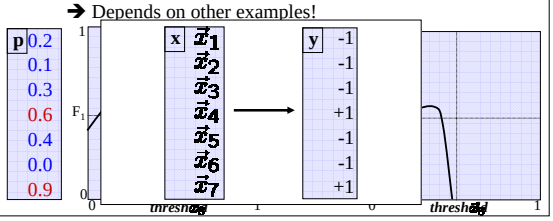


Examples of Complex Output Spaces

- **Non-Standard Performance Measures (e.g. F_1-score, Lift)**
  - F_1-score: harmonic average of precision and recall

$$F_1 = \frac{2\,Prec\,Rec}{Prec + Rec}$$

  - New example vector $\vec{x}_8$. Predict $y_8=1$, if $P(y_8=1|\vec{x}_8)=0.4$?
    - ➔ Depends on other examples!



(Recent) Developments

- **Generative Models**
  - Hidden Markov models
  - Probabilistic context-free grammars
  - Pair HMM alignment models
  - Bayesian networks
  - Etc.
- **Discriminative Models**
  - Perceptron training of HMM
  - Conditional random fields
  - Structural SVMs
  - Maximum-margin Markov networks
  - Kernel dependency estimation
  - Conditional pseudo likelihood
  - Transformer networks
  - Etc.

What will we do in this class?
- Overview of discriminative methods for structured outputs
- Understand strength and weaknesses compared to conventional methods
- Relationship between methods
- Directions for research

Course Overview

- **Meetings (Tuesdays, 1:25pm – 2:40pm)**
  - Background lectures:
    - Few lectures to provide background
  - Discussion of research papers:
    - Student presents paper in class
    - Student leads discussion of paper
- **Homework Assignments**
  - Read papers for next week
  - Prepare questions and for discussion
- **Project**
  - Optional
- **Exams**
  - None

Paper Presentations

- **Selection**
  - Vote for papers you like
- **Content**
  - Self-contained explanation of paper
  - Might need to add background
  - Add examples to make paper more clear
  - Prepare slides
- **Preparation / Feedback**
  - I am happy to help with making sense of the material
  - By Friday before presentation, go over slides with me
  - I will give feedback after the presentation

Ways to Take the Class

- **Enrolling for "Pass/Fail":**
  - Present (multiple) papers in class
  - Lead the discussion during this class
  - Read all other papers discussed in class
  - Participate in discussion of other papers
  - Take in-class quizzes
  - Grading: presentation, participation, quizzes
- **Enrolling for "Letter Grade":**
  - All of the above
  - Do self-defined project
  - Grading: project, presentation, participation, quizzes
- **Enrolling as "Audit":**
  - No

Project (Optional)

- **Topic of project is self-defined**
  - Related to your research interests
  - Related to predicting structured outputs
  - Can be an application to a particular problem, comparison of methods, development of new methods, etc.
- **Schedule**
  - Project proposal (draft within 2 weeks)
  - Regular status reports
  - Final project report (due during week of finals)

Prerequisites

- **Basic understanding of machine learning**
  - CS 478 or CS 578 or other machine learning class
    - CS 472 or other general AI introduction (borderline)
- **Basic linear algebra**
- **Basic probability**

- **Related courses you might consider instead of CS778**
  - CS472: Foundations of Artificial Intelligence
  - CS578: Empirical Methods in Machine Learning

Selecting Your Papers

- **Today:**
  - Look at handout or "reference" section on course homepage http://www.cs.cornell.edu/Courses/cs778/2006fa/
  - Find papers that you are interested in presenting
  - Did I forget any interesting papers you would like to read?
- **Tomorrow (Wednesday, Aug. 30) by 8pm:**
  - Send me via email the paper you would be interested in presenting
    - at least four papers marked with "***" (ranked, if you want)
    - as many other papers as you like (ranked, if you want)
- **Thursday, Aug. 31:**
  - I will send around the paper assignments and schedule

How to Get in Touch

- **Lecture**
  - Tuesdays, 1:25pm - 2:40pm, Upson 5126
- **WWW Page**
  - http://www.cs.cornell.edu/Courses/cs778/2006fa/
- **Email**
  - Thorsten Joachims: tj@cs.cornell.edu
  - Mailing list: TBA
- **Office Hours**
  - Thorsten Joachims: Mondays at 1:00pm – 2:00pm, 4153 Upson Hall