

# Learning and Inference in Structured Prediction

Final words



# Structured Prediction

- **Examples**

- Named entities + Relations
- Computer Vision
- Additional NLP Examples

- **Problem Formulation**

- Constrained Conditional Models: Integer Linear Programming Formulations

- **Constraints Driven Learning**

# Structured output is...

- A **graph**, possibly labeled and/or directed
  - Possibly from a restricted family, such as chains, trees, etc.
  - A discrete representation of input
  - Eg. A table, the SRL frame output, a sequence of labels etc
- A collection of **inter-dependent decisions**
  - Eg: The sequence of decisions used to construct the output
- The result of a **combinatorial optimization** problem
  - $\operatorname{argmax}_{\mathbf{y} \in \text{all outputs}} \text{score}(\mathbf{x}, \mathbf{y})$

Representation

Procedural

Formally

# How did we get here?

## Binary classification

- Learning algorithms
- Prediction is easy:  
Threshold
- Features (???)



## Multiclass classification

- Different strategies
  - One-vs-all, all-vs-all
- Global learning algorithms
- One feature vector per outcome
  - Each outcome scored
- Prediction = highest scoring outcome



## Structured classification

- Global models or local models
- Each outcome scored
- Prediction = highest scoring outcome
- **Inference is no longer easy!**
  - **Makes all the difference**

# Learning a Structured Prediction Model

- **Local Learning v.s. Global Learning**
- **Global Learning Algorithms**
  - Online learning: Structured Perceptron
  - Batch learning: Structured SVM
- **Optimization methods for Structured SVM**
  - Stochastic Gradient Decent
  - Dual Coordinate Descent
  - Learning on a multi-core machine

# Amortized ILP Inference

- Taking advantage of regularities in the output space to speed up prediction
- **Amortization at Inference Time**
  - Theorems
  - Decomposition
- **Amortization during Learning**
  - Approximate Inference

# Distributed Representations & Structures

- ***Outputs*** are discrete objects
  - One of a set of labels (document classification)
  - Label sequences (POS tagging, Chunking, NER)
  - Trees with labeled edges/nodes (Parsing)
  - Arbitrary graphs (Semantic Role Labeling, event extraction)
- ***Distributed representations for structures***
  - Starting with individual labels to compose full structures
  - A natural generalization of standard structured prediction formalism

# Structured Prediction Software

- *Illinois Structured Learning Library (IllinoisSL)*
  - Starting with individual labels to compose full structures
  - A natural generalization of standard structured prediction formalisms

`http://cogcomp.cs.illinois.edu/software/illinois-sl`

# Looking ahead

- **Representations?**

- Can we learn the factorization of the output into parts?
- Can we learn feature functions?

- **Supervision?**

- Clever tricks to get data
- Using latent variable learning

- **Declarative programming languages for structured outputs**

- (DeLBP workshop right here)

- **Applications?**

- How does structured prediction help you?
- Gathering importance as computer programs have to deal with uncertain, noisy inputs and make **complex,**  
**interconnected** decisions