CPSC 540: Machine Learning
Structured Prediction Motivation

Mark Schmidt
University of British Columbia
Winter 2020
Motivation: Structured Prediction

Classic **supervised learning** focuses on predicting single discrete/continuous label:

**Input:** \( P \)

**Output:** "P"

**Structured prediction** allows **general objects** as labels:

**Input:** \( Paris \)

**Output:** "Paris"
Two ways to formulate as “classic” machine learning:

1. Treat each word as a different class label.
   - Problem: there are too many possible words (huge numbers “coupons” to collect).
   - You will never recognize new words.

2. Predict each letter individually:
   - Works if you are really good at predicting individual letters.
   - But some tasks don’t have a natural decomposition.
   - Ignores dependencies between letters.
Motivation: Structured Prediction

- What letter is this?

- What are these letters?

- Predict each letter using “classic” ML and features from neighbouring images?
  - Can be good or bad depending on goal:
    - Good if you want to predict individual letters.
    - Bad if goal is to predict entire word.
Examples of Structured Prediction

Translate

I moved to Canada in 2013, as indicated on my 2013 declaration of revenue. I received no income from French sources in 2014. How can I owe 12 thousand Euros?

Je déménagé au Canada en 2013, comme indiqué sur ma déclaration de revenu 2013. Je recevais aucun revenu de source française en 2014. Comment puis-je dois 12 mille euros?

```
S
  / \     /
 NP  VP   /
  /   /  /
 Det N  V  NP
 /  \
 The teacher praised Det N
   \   /
    the student
```
Examples of Structured Prediction

Coding Regions

Non-coding Regions
(Containing large TE content)
Examples of Structured Prediction
Does the brain do structured prediction?

Does the brain do structured prediction?

Gestalt effect: “whole is other than the sum of the parts”.

What do you see?
By shifting perspective you might see an old woman or a young woman.
Summary

- “How much data do you need” question and coupon collecting.
- Stochastic gradient descent on the training error or test error (with one pass).
- $O(1/n)$ error rate when you have $n$ training examples.
- Structured prediction: supervised learning with complicated “labels”.

- Next time: everyone’s favourite distributions...