Lecture 7

Predicting Structured Outputs

Part 1: Interlude: What Are Structured Outputs

Machine Learning for Structured Data

Vlad Niculae · LTL, UvA · `https://vene.ro/mlsd`

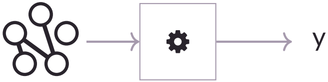
Predicting Structured Outputs

1. **Interlude: What Are Structured Outputs**

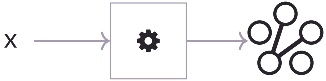
2. Probabilistic Models of Structures

So far, we've studied this scenario:

- Structured inputs
- Familiar unstructured outputs: classification / regression.

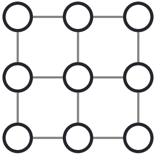


In the next part of class,
we study **structured outputs**.

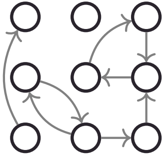


Reminder: Kinds of Structure

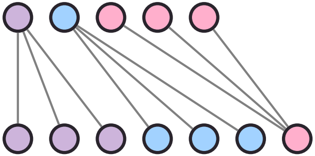
Sequence



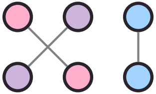
Grid



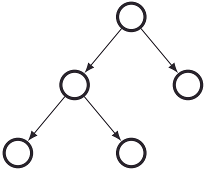
Graph



Alignments



Permutations



Hierarchy

Structured outputs are:

- discrete objects
- made of smaller parts
- which interact with each other and/or constrain each other.

Structured outputs are:

- discrete objects
- made of smaller parts
- which interact with each other and/or constrain each other.

Example: What are the possible ways to assign 4 jockeys to 4 horses?

$$\begin{aligned}\mathcal{Y} = \{&(1, 2, 3, 4),\\ &(1, 2, 4, 3),\\ &(1, 3, 2, 4),\\ &\ldots,\\ &(4, 3, 2, 1)\}\end{aligned}$$

Structured outputs are:

- discrete objects
- made of smaller parts
- which interact with each other and/or constrain each other.

Example: What are the possible ways to assign 4 jockeys to 4 horses?

$$\begin{aligned}\mathcal{Y} = \{&(1,2,3,4),\\ &(1,2,4,3),\\ &(1,3,2,4),\\ &\ldots,\\ &(4,3,2,1)\}\end{aligned}$$

We can't just predict the best jockey for each horse, or the best horse for each jockey, since we might end up with double assignments.

Structured outputs are:

- discrete objects
- made of smaller parts
- which interact with each other and/or constrain each other.

Example: What are the possible ways to assign 4 jockeys to 4 horses?

$$\begin{aligned}\mathcal{Y} = \{&(1,2,3,4),\\ &(1,2,4,3),\\ &(1,3,2,4),\\ &\ldots,\\ &(4,3,2,1)\}\end{aligned}$$

We can't just predict the best jockey for each horse, or the best horse for each jockey, since we might end up with double assignments.

What is $|\mathcal{Y}|$?

Recap: Logistic Regression and Perceptron Losses

The two losses we've seen for multi-class classification:
(changing notation slightly)

$$L_{\mathrm{LR}}(y) = -\log \Pr(Y = y|x) \qquad = -\,\mathrm{score}(y) + \log \sum_{y' \in \mathcal{Y}} \exp\left(\mathrm{score}(y')\right)$$

$$L_{\mathrm{Perc}}(y) \qquad = -\,\mathrm{score}(y) + \max_{y' \in \mathcal{Y}} \mathrm{score}(y')$$

For classification:

- we had $\mathcal{Y} = \{1, 2, \ldots, K\}$
- the model (linear or NN) outputs a vector $\boldsymbol{a}$ of scores for each class, so $\mathrm{score}(y) = a_y$.

Can we generalize this to structured $\mathcal{Y}$?

Lecture 7

Predicting Structured Outputs

Part 2: Probabilistic Models of Structures

Machine Learning for Structured Data

Vlad Niculae · LTL, UvA · `https://vene.ro/mlsd`

Predicting Structured Outputs

1. Interlude: What Are Structured Outputs

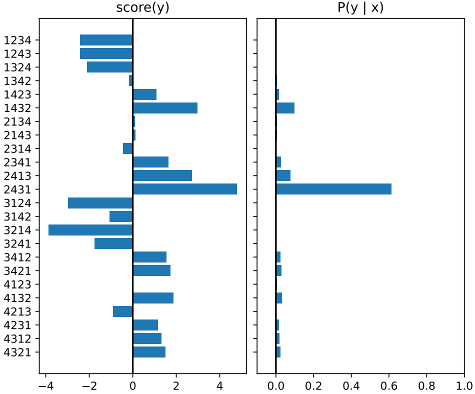
2. Probabilistic Models of Structures

Probabilistic Models of Structures

Our model must be able to assign a score to every possible structure, $\mathrm{score}(y; x, \theta)$. For brevity we just write $\mathrm{score}(y)$, but remember it depends on input and params.

From this, we can get a probability distribution over possible structures:

$$\Pr(y \mid x) = \frac{\exp\left(\mathrm{score}(y)\right)}{\sum_{y' \in \mathcal{Y}} \exp\left(\mathrm{score}(y')\right)}$$



Modelling challenges

Essential computational prerequisites:

- $\text{score}(y)$
- for prediction: $\arg\max_{y \in \mathcal{Y}} \text{score}(y)$
- for learning: $\log \sum_{y \in \mathcal{Y}} \exp\left(\text{score}(y)\right)$

The challenges: unlike multi-class classification,

- $\mathcal{Y}$ can vary for each data point (e.g., with n. horses)
- $|\mathcal{Y}|$ can get very large: we can't just for-loop over it.

Generally intractable!
But, for certain structures and scoring functions, efficient algorithms exist.

The Road Ahead

In the rest of the class, we shall cover a wide range of structured output tasks:

- Sequence labelling
- Sequence segmentation
- Alignments between sequences;
- Assignments and permutations
- Grid / graph labelling

While there is no general-purpose structure prediction algorithm, we shall learn three main tools that will get you far:

- Dynamic programming
- Integer linear programming
- Gibbs sampling