1 Midterm exam

1.1 Question 1

Consider the following input text:

*Jose Saramago was born in 1922 into a family of landless peasants in Azinhaga, Portugal. In 1924, Saramago’s family moved to Lisbon, where his father started working as a policeman.*

1. Show the text segmented into sentences, where each sentence is further segmented into tokens.

2. Show what the coreference resolution task is supposed to find in this text.

3. What NLP task is responsible for determining that ‘family’ is a noun and ‘moved’ is a verb at past tense?

4. What NLP task is responsible for determining that the noun ‘family’ is the subject of the verb ‘moved’?

Answer:

1. $S_1 = \text{Jose.Saramago.was.born.in.1922.into.a.family.of.landless.peasants.in.Azinhaga.Portugal.}$
   $S_2 = \text{In.1924,.Saramago.'s.family.moved.to.Lisbon,.where.his.father.}
   \text{started.working.as.a.policeman.}$

2. \{Jose.Saramago$\text{1}$, Saramago$\text{2}$, his$\text{2}$\}
   \{a.family.of.landless.peasants$\text{1}$, Saramago's.family$\text{2}$\}
   \{Lisbon, where\}
   \{his.father, a.policeman\}

3. POS tagging.

4. Syntactic parsing.

1.2 Question 2

See answer in `midterm.py`. 
1.3 Question 3
1.4 Question 4

Given the following short emails, each labeled with either as spam or ham:

E1 = million, award, claim, claim [spam]
E2 = work, today, project, send, project, [ham]
E3 = award, send, work, million, million [spam]
E4 = today, project, project, million, claim, work [ham]

and a new document:

D = work, award, project, send

compute the most likely class for D. Assume a naive Bayes classifier and use add-1 smoothing for the likelihoods. Show your work.

Vocabulary is $V = \{\text{award, claim, million, project, send, today, work}\}$ with size 7.

$P(\text{ham}) = 2/4 = 0.5$

$P(\text{work}|\text{ham}) = \frac{2 + 1}{11 + 7} = \frac{3}{18}$

$P(\text{award}|\text{ham}) = \frac{0 + 1}{11 + 7} = \frac{1}{18}$

$P(\text{project}|\text{ham}) = \frac{4 + 1}{11 + 7} = \frac{5}{18}$

$P(\text{send}|\text{ham}) = \frac{1 + 1}{11 + 7} = \frac{2}{18}$

$P(\text{ham}|D) \sim 0.5 \times \frac{3 \times 1 \times 5 \times 2}{18^4} = \frac{15}{18^4} = 0.000142$

$P(\text{spam}) = 2/4 = 0.5$

$P(\text{work}|\text{spam}) = \frac{1 + 1}{9 + 7} = \frac{2}{16}$

$P(\text{award}|\text{spam}) = \frac{2 + 1}{9 + 7} = \frac{3}{16}$

$P(\text{project}|\text{spam}) = \frac{0 + 1}{9 + 7} = \frac{1}{16}$

$P(\text{send}|\text{spam}) = \frac{1 + 1}{9 + 7} = \frac{2}{16}$

$P(\text{spam}|D) \sim 0.5 \times \frac{2 \times 3 \times 1 \times 2}{16^4} = \frac{6}{16^4} = 0.000091$

We computed that $P(\text{ham}|D) > P(\text{spam}|D)$, therefore the most likely class is ham.
1.5 HHW02: Theory exercise 4.1

The class posterior probabilities are as follows:

- \( P(\text{pos}|D) = 0.5 \times 0.09 \times 0.07 \times 0.29 \times 0.04 \times 0.08 = 2.92 \times 10^{-6} \)
- \( P(\text{neg}|D) = 0.5 \times 0.16 \times 0.06 \times 0.06 \times 0.15 \times 0.11 = 4.75 \times 10^{-6} \)

Therefore, the predicted class is \textit{negative}.

1.6 HHW02: Theory exercise 4.2

Vocabulary is \( V = \{\text{couple, fast, fly, fun, furious, love, shoot}\} \) with size 7.

- \( P(\text{com}) = \frac{2}{5} = 0.4 \)
- \( P(\text{couple}|\text{com}) = \frac{2 + 1}{9 + 7} = \frac{3}{16} \)
- \( P(\text{fast}|\text{com}) = \frac{1 + 1}{9 + 7} = \frac{2}{16} \)
- \( P(\text{fly}|\text{com}) = \frac{1 + 1}{9 + 7} = \frac{2}{16} \)
- \( P(\text{shoot}|\text{com}) = \frac{0 + 1}{9 + 7} = \frac{1}{16} \)
- \( P(\text{com}|D) \sim 0.4 \times \frac{3 \times 2 \times 2 \times 1}{16^4} = \frac{4.8}{16^4} = 73.2 \times 10^{-6} \)
- \( P(\text{com}) = \frac{3}{5} = 0.6 \)
- \( P(\text{couple}|\text{act}) = \frac{0 + 1}{11 + 7} = \frac{1}{18} \)
- \( P(\text{fast}|\text{act}) = \frac{2 + 1}{11 + 7} = \frac{3}{18} \)
- \( P(\text{fly}|\text{act}) = \frac{1 + 1}{11 + 7} = \frac{2}{18} \)
- \( P(\text{shoot}|\text{act}) = \frac{4 + 1}{11 + 7} = \frac{5}{18} \)
- \( P(\text{act}|D) \sim 0.6 \times \frac{1 \times 3 \times 2 \times 5}{18^4} = \frac{18}{18^4} = 171.4 \times 10^{-6} \)

We computed that \( P(\text{act}|D) > P(\text{com}|D) \), therefore the most likely class is \textit{action}. 
We have a vector \( z = [z_1, z_2, ..., z_K] \).

\[
z_k' = \frac{z_k}{\sum_{j=1}^{K} z_j} \quad (1)
\]

We have that \( \sum_{k=1}^{K} z_k' = 1 \)

Let \( K = 2 \) and \( z_1 = -1 \) and \( z_2 = +3 \)

How about this:

\[
z_k' = \frac{|z_k|}{\sum_{j=1}^{K} |z_j|} \quad (2)
\]

This has the issue that it creates the same 'probability' \( z_k' \) for \( z_k = -1 \) vs. \( z_k = +1 \).

How about this? This is the softmax transformation.

\[
p_k = z_k' = \frac{e^{z_k}}{\sum_{j=1}^{K} e^{z_j}} \quad (3)
\]

\[
p_k = z_k' = \frac{\exp(z_k)}{\sum_{j=1}^{K} \exp(z_j)} \quad (4)
\]

where the probability vector is \( p = p_1, p_2, ..., p_K \). The vectorized formulation is:

\[
p = softmax(z) \quad (5)
\]

At test time, we do not need to compute the denominator, or the exponentials:

\[
\hat{c}(x) = \arg \max_{1 \leq k \leq K} w_k^T x + b_k \quad (6)
\]