Margot Wagner

A53279875

CSE 250a HW4

Due: 11/3/20

Text, letter

Description automatically generated

Text, letter

Description automatically generated

Text, letter

Description automatically generated

Text

Description automatically generated

4.3 Statistical Language Processing

(a) Start with the letter “A”

Table

Description automatically generated

(b) 5 most likely words to follow “THE”

A picture containing table

Description automatically generated

(c) “Last week the stock market fell by one hundred points”

Log-likelihood of unigram model: **-41.643**

Log-likelihood of bigram model: **-44.740**

The **bigram model** yields the highest log-likelihood.

(d) “The nineteen officials sold fire insurance”

Log-likelihood of unigram model: **-41.643**

Log-likelihood of bigram model: **undefined**

**“Nineteen officials”** and **“sold fire**” are not seen in the training corpus. As a result, these **probabilities are zero**, bringing the entire probability to zero. The log likelihood of zero is undefined as e approaches negative infinity for values approaching zero.

(e) Mixture model

A picture containing chart

Description automatically generated

The optimal lambda is 0.41.

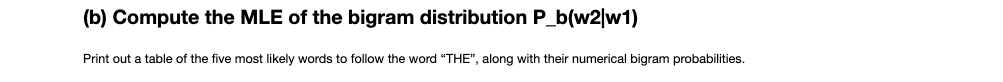
(f)

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application

Description automatically generated



Graphical user interface, text

Description automatically generated

Graphical user interface

Description automatically generated

Graphical user interface, application

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Graphical user interface, chart

Description automatically generated

**4.4 Stock Market prediction**

(a) Linear coefficients

**a1 = 0.9507**

**a2 = 0.0156**

**a3 = 0.0319**

(b) Mean squared prediction error

**MSE2000 = 13902.40**

**MSE2001 = 2985.10**

These MSEs are very large, so I **would not recommend** this model for stock market predictions.

(c) Source code

A picture containing table

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated