Prof. Dr. Andreas Hotho,
M.Sc. Janna Omeliyanenko
Lecture Chair X for Data Science, Universität Würzburg

4. Exercise for "Sprachverarbeitung und Text Mining"

26.11.2021

1 Knowledge Questions

1. When is a Maximum Entropy Model called consistent to the given data? Describe the definition with your own words.

> Consistency condition: $\hat{E}(f_i) = E(f_i)$
> The conditional probability $p(y|x)$ needs to be trained such that for every feature f_i the expected value $E(f_i)$ can reproduce the expected value $\hat{E}(f_i)$ observed within the data.

2. Name the three conditions that must be enforced during optimization of a Maximum Entropy Model. Describe their purpose in your own words.

> $$\hat{E}(f_i) - E(f_i) = 0, \forall f_i \tag{1}$$
> Classifier needs to match our training data
> $$\sum_{y \in Y} p(y|x) = 1, \forall x \tag{2}$$
> $$p(y|x) \geq 0, \forall x, y \tag{3}$$
> We want $p(y|x)$ to be a real probability distribution.

3. When employing Maximum Markov Models in practice, how are the different features determined that are used within the model?

Options that can be considered:

- Construct templates that can mine features from data
  (e.g. current-word $\wedge$ current-tag, current-word $\wedge$ previous-tag $\wedge$ current-tag, ...)
- Construct single features through expert knowledge
- Import known useful features (known linguistic rules, dictionaries, ...)

2 Maximum Entropy Models

Consider the sentence “Wir rennen oft zum Bus”. (English: “We often run to the bus”.)

Classify the word “rennen” (English: “run”) using the following Maximum Entropy Model.

Feature	λ for PPER	λ for VVFIN	λ for ADV	λ for APPRART	λ for NN
$x_i = Wir$	0.9	-0.2	0.1	0.1	0.2
$x_i = rennen$	0.2	0.8	0.15	-0.11	0.45
$x_i = oft$	0.1	0.05	0.9	0.15	0.01
$x_i = zum$	0.01	0.05	0.2	0.8	0.15
$x_i = Bus$	0.1	0.02	0.1	0.2	0.9
$x_{i-1} = Wir$	0.02	0.98	0.2	0.3	0.03

$D = \exp(0.2{+}0.02){+}\exp(0.8{+}0.98){+}\exp(0.15{+}0.2){+}\exp(-0.11{+}0.3){+}\exp(0.45{+}0.03)$

$$P(PPER|x) = \frac{\exp(0.2+0.02)}{D} = 0.1094$$

$$P(VVFIN|x) = \frac{\exp(0.8+0.98)}{D} = \mathbf{0.5190}$$

$$P(ADV|x) = \frac{\exp(0.15+0.2)}{D} = 0.1242$$

$$P(APPRART|x) = \frac{\exp(-0.11+0.3)}{D} = 0.1058$$

$$P(NN|x) = \frac{\exp(0.45+0.03)}{D} = 0.1414$$

$\Rightarrow$ “rennen“ is classified as VVFIN.

3 Lagrangians

Solve the optimization problem that is defined by the function

$$f(x, y) = x^2 - 2y$$

and the condition

$$g(x, y) = x + y - 5 = 0$$

by using the method of Lagrange Multipliers. Did you find a maximum or a minimum?

$$L(x, y, \lambda) = f(x, y) = x^2 - 2y - \lambda \cdot (x + y - 5) \tag{4}$$

Calculating the Gradient:

$$\Delta L(x, y, \lambda) = \begin{pmatrix} 2x - \lambda \\ -2 - \lambda \\ -x - y + 5 \end{pmatrix} \tag{5}$$

So we get the following system of equations with three unkown varibles:

$$\begin{array}{rr} \textbf{(I)} & 2x - \lambda = 0 \\ \textbf{(II)} & -2 - \lambda = 0 \\ \textbf{(III)} & x + y - 5 = 0 \end{array}$$

We can solve this and get the values $x = -1, y = 6, \lambda = -2$ We found a minimum, since the function describes an upwardly opened parabola.