Máster Universitario en Ingeniería Informática

Sistemas Inteligentes

Unit 6. Maximum Entropy Models - Exercises

2022/2023

Outline

Outline

1. Exercises

2. Solutions

Exercise 1

Given a set of words, such that each word has associated a class label c_0 or c_1, $C = \{(w_0, c_0), (w_0, c_0), (w_1, c_0), (w_1, c_0), (w_1, c_0), (w_1, c_0), (w_0, c_1), (w_2, c_1), (w_2, c_1), (w_2, c_1)\}$, it is desired to obtain a maximum entropy model for classifying texts that include these words. The following features have been defined for this purpose:

$$f(x, y) = \begin{cases} 1 & \text{if } y = c_i \text{ and it has the word } x = w_j \\ 0 & \text{otherwise} \end{cases}$$

1. Suppose that all λ_i are initialized to 0, then compute the increasing δ_0 for the feature $f_0 = f(w_0, c_0)$ with the IIS algorithm

2. Suppose that a ME model has been learned and the following values have been obtained:

$$\begin{aligned} \lambda_0 &= 0.3 \text{ associated to the feature } f_0 = f(w_0, c_0) \\ \lambda_1 &= 3.0 \text{ associated to the feature } f_1 = f(w_1, c_0) \\ \lambda_2 &= -0.4 \text{ associated to the feature } f_2 = f(w_0, c_1) \\ \lambda_3 &= 3.0 \text{ associated to the feature } f_3 = f(w_2, c_1) \end{aligned}$$

   Compute in which class is classified a text that includes the word w_0.

Exercise 2

The gender of a bird species can be known from three features a_0, a_1 and a_2. We have the following set of samples:

$$C = \{((a_0, a_1), M), ((a_0, a_1), M), ((a_1, a_2), M),$$
$$((a_0, a_2), F), ((a_0, a_2), F), ((a_0, a_2), F)\}$$

where M indicates that the bird is a male and F indicates that it is a female. To obtain a ME model for classifying the birds, the following features have been defined:

$$f(x, y) = \begin{cases} 1 & \text{if } y = S \text{ and the feature } a_j \text{ is in } x \\ 0 & \text{otherwise} \end{cases}$$

where $S \in \{M, F\}$.

1. Suppose that all λ_i are initialized to 0, then compute the increasing δ for the feature $f = f(a_0, F)$ with the IIS algorithm.
2. Suppose that a ME model has been learned and the following values have been obtained:

$$\begin{array}{llllll} \lambda_{a_0,M} = & -0.157 & \lambda_{a_1,M} = & 1.418 & \lambda_{a_2,M} = & -0.801 \\ \lambda_{a_0,F} = & 0.122 & \lambda_{a_1,F} = & 0.0 & \lambda_{a_2,F} = & 0.385 \end{array}$$

   Compute the most probable class of a bird if the features a_1 and a_2 have been observed.

Exercise 3

Let be a classification problem into three classes A, B and C, where the classification is carried out with a Maximum Entropy classifier. The classification of each sample is performed according to 3 features out of 5, noted as $(a_0, a_1, a_2, a_3, a_4)$. The features of the Maximum Entropy classifier are defined as:

$$f(x, y) = \begin{cases} 1 & \text{if } y = S \text{ and the feature } a_j \text{ is active in } x \\ 0 & \text{otherwise} \end{cases}$$

where $S \in \{A, B, C\}$

Supose that a model has been trained and the obtained parameters λ_{fc} (where $f \in \{a_0, \ldots, a_4\}$, $c \in \{A, B, C\}$) are:

	Feature				
Class	a_0	a_1	a_2	a_3	a_4
A	0.37	0.0	-0.04	0.08	0.0
B	0.0	0.0	-0.04	-0.05	-0.28
C	0.0	0.23	0.06	-0.05	0.32

Compute the class of a sample if the features a_0, a_2 and a_3 have been observed.

Exercise 4

We have a classification problem with classes c_0 and c_1. Each sample is a string with three symbols, $s = s_0 s_1 s_2$, such that $s_i \in \{a, b\}$. A Maximum Entropy model is defined for the problem, with features:

$$f(x, y) = \begin{cases} 1 & \text{if } y = C \text{ and } x = t_i \text{ means that symbol } t \text{ is in position } i \text{ of string } s \\ 0 & \text{otherwise} \end{cases}$$

Where $C = \{c_0, c_1\}$ and $t \in \{a, b\}$.

The model has been estimated with the following parameters:

Class 0

	Position		
Symbol	0	1	2
a	0.0	0.096	-0.074
b	0.170	-0.051	0.061

Class 1

	Position		
Symbol	0	1	2
a	0.231	-0.135	0.061
b	-0.366	0.045	-0.074

Classify the string “abb” according the model.

Outline

1. Exercises

2. Solutions

Exercise 1

1. The posterior probability is:

$$p(y|x) = \frac{1}{Z(x)} \exp\left(\sum_i \lambda_i f_i(x,y)\right) \qquad Z(x) = \sum_y \exp\left(\sum_i \lambda_i f_i(x,y)\right)$$

We have to calculate $\delta_0 = \frac{1}{M} \log \frac{\tilde{p}(f_0)}{p_\lambda(f_0)}$, where $\tilde{p}(f_0) = \sum_{x,y} \tilde{p}(x,y) f_0(x,y)$ and $p_\lambda(f_0) = \sum_{x,y} \tilde{p}(x) p_\lambda(y|x) f_0(x,y)$
Thus:

$$\delta_0 = \frac{1}{M} \log \frac{\sum_{x,y} \tilde{p}(x,y) f_0(x,y)}{\sum_{x,y} \tilde{p}(x) p_\lambda(y|x) f_0(x,y)} = \frac{1}{M} \log \frac{\sum_{x,y} \tilde{p}(x,y)}{\sum_{x,y} \tilde{p}(x) p_\lambda(y|x)}$$

Since it is specifically for (w_0, c_0), $\tilde{p}(x,y) = \tilde{p}(w_0, c_0) = \frac{2}{10} = \frac{1}{5}$, $\tilde{p}(x) = \tilde{p}(w_0) = \frac{3}{10}$, $p_\lambda(y|x) = \frac{\exp 0}{\sum_{y \in \{c_0, c_1\}} \exp 0} = \frac{1}{2}$, and $M = f^{\#}(x,y) = f^{\#}(w_0, c_0) = 1$. Then:

$$\delta_0 = \frac{1}{1} \log \frac{\frac{1}{5}}{\frac{3}{10}\frac{1}{2}} = \log \frac{4}{3} \approx 0.2877$$

Exercise 1

2. Since $x = w_0$ and $y \in \{c_0, c_1\}$, we have to calculate $p(c_0|w_0)$ and $p(c_1|w_0)$; in general:

$$p(y|x) = \frac{1}{Z(x)} \exp\left(\sum_i \lambda_i f_i(x,y)\right) \qquad Z(x) = \sum_y \exp\left(\sum_i \lambda_i f_i(x,y)\right)$$

Thus, we must calculate $Z(x) = Z(w_0)$:

$$Z(w_0) = \sum_{\{c_0,c_1\}} \exp\left(\sum_i \lambda_i f_i(w_0,y)\right) = \exp\left(\sum_i \lambda_i f_i(w_0,c_0)\right) +$$

$$\exp\left(\sum_i \lambda_i f_i(w_0,c_1)\right) = \exp \lambda_0 f_0(w_0,c_0) + \exp \lambda_2 f_2(w_0,c_1) = \exp 0.3 + \exp -0.4$$

Consequently, $p(c_0|w_0) = \frac{1}{Z(w_0)} \exp\left(\sum_i \lambda_i f_i(w_0,c_0)\right) = \frac{1}{Z(w_0)} \exp \lambda_0 f_0(w_0,c_0) = \frac{\exp 0.3}{\exp 0.3 + \exp -0.4} \approx 0.668$, and $p(c_1|w_0) = \frac{1}{Z(w_0)} \exp\left(\sum_i \lambda_i f_i(w_0,c_1)\right) = \frac{1}{Z(w_0)} \exp \lambda_2 f_2(w_0,c_1) = \frac{\exp -0.4}{\exp 0.3 + \exp -0.4} \approx 0.3318$, and $c(w_0) = c_0$.

Exercise 2

1. Taking into account that $\lambda_i = 0$ and $f = f(F, c_0)$, the increment δ corresponds to $\delta = \frac{1}{M} \log \frac{\tilde{p}(f)}{p_\lambda(f)}$. Since:

$$\tilde{p}(f) = \sum_{x,y} \tilde{p}(x,y) f(x,y) = \tilde{p}(a_0, F) f(a_0, F)$$

$$p_\lambda(f) = \sum_{x,y} \tilde{p}(x) p_\lambda(y|x) f(x,y) = \tilde{p}(a_0) p_\lambda(F|a_0) f(a_0, F)$$

Thus, $\delta = \frac{1}{M} \log \frac{\tilde{p}(a_0,F)}{\tilde{p}(a_0) p_\lambda(F|a_0)}$
Since a_0 is present in the sample for (a_0, a_1) and (a_0, a_2),
$M = f^{\#}(x,y) = f^{\#}(a_0, F) = 2$. Now, as $\tilde{p}(a_0, F) = \frac{3}{6} = \frac{1}{2}$ and $\tilde{p}(a_0) = \frac{5}{6}$,
$p_\lambda(F|a_0)$ must be calculated:

$$p_\lambda(F|a_0) = \frac{1}{Z(a_0)} \exp\left(\sum_i \lambda_i f(a_0, F)\right) \quad Z(a_0) = \sum_{y \in \{M,F\}} \exp\left(\sum_i \lambda_i f(a_0, F)\right)$$

Since $Z(a_0) = 2$, $p_\lambda(F|a_0) = \frac{1}{2}$, and finally
$\delta = \frac{1}{2} \log \frac{\frac{1}{2}}{\frac{5}{6}\frac{1}{2}} = \frac{1}{2} \log \frac{5}{6} \approx 0.091$

Exercise 2

2. Taking into account λ values and that $x = (a_1, a_2)$:

$$p(y|(a_1, a_2)) = \frac{1}{Z((a_1, a_2))} \exp \sum_i \lambda_i f(x, y)$$

Then, we calculate initially $Z((a_1, a_2))$ taking into account that $f(a_0, y) = 0$ and $f(a_1, y) = f(a_2, y) = 1$:

$$Z((a_1, a_2)) = \sum_{y \in \{M,F\}} \exp\left(\sum_i \lambda_i f_i(x, y)\right) = \sum_{y \in \{M,F\}} \exp(\lambda_{a_1,y} + \lambda_{a_2,y}) =$$

$$\exp(\lambda_{a_1,M} + \lambda_{a_2,M}) + \exp(\lambda_{a_1,F} + \lambda_{a_2,F}) = \exp 0.617 + \exp 0.385$$

Thus, $p(M|(a_1, a_2)) = \frac{\exp 0.617}{\exp 0.617 + \exp 0.385} \approx 0.55$, and $p(F|(a_1, a_2)) = \frac{\exp 0.385}{\exp 0.617 + \exp 0.385} \approx 0.44$; consequently, $c(x) = M$

Exercise 3

As usual:

$$p(y|x) = \frac{1}{Z(x)} \exp\left(\sum_i \lambda_i f_i(x, y)\right) \qquad Z(x) = \sum_y \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

Since $x = (a_0, a_2, a_3)$, we calculate $Z(x)$ taking into account that $f(a_0, y) = f(a_2, y) = f(a_3, y) = 1$ and $f(a_1, y) = f(a_4, y) = 0$. That is:

$$Z((a_0, a_2, a_3)) = \sum_{y\in\{A,B,C\}} exp\left(\sum_i \lambda_i f_i(x, y)\right) = \sum_{y\in\{A,B,C\}} \exp(\lambda_{a_0,y}+\lambda_{a_2,y}+\lambda_{a_3,y}) =$$

$$\exp(\lambda_{a_0,A}+\lambda_{a_2,A}+\lambda_{a_3,A})+\exp(\lambda_{a_0,B}+\lambda_{a_2,B}+\lambda_{a_3,B})+\exp(\lambda_{a_0,C}+\lambda_{a_2,C}+\lambda_{a_3,C}) =$$

$$\exp(0.41) + \exp(-0.09) + \exp(0.01)$$

Then, $p(A|x) = \frac{\exp(0.41)}{\exp(0.41)+\exp(-0.09)+\exp(0.01)} \approx 0.44$,
$p(B|x) = \frac{\exp(-0.09)}{\exp(0.41)+\exp(-0.09)+\exp(0.01)} \approx 0.27$, and
$p(C|x) = \frac{\exp(0.01)}{\exp(0.41)+\exp(-0.09)+\exp(0.01)} \approx 0.29$. Consequently, $c(x) = A$.

Exercise 4

As usual:

$$p(y|x) = \frac{1}{Z(x)} \exp\left(\sum_i \lambda_i f_i(x,y)\right) \qquad Z(x) = \sum_y \exp\left(\sum_i \lambda_i f_i(x,y)\right)$$

Since s="abb", $f(a_0, y) = f(b_1, y) = f(b_2, y) = 1$, and the rest are 0. Thus:

$$Z(x) = \sum_{y \in \{c_0, c_1\}} \exp(\lambda_{a_0,y} + \lambda_{b_1,y} + \lambda_{b_2,y}) =$$

$$\exp(\lambda_{a_0,c_0} + \lambda_{b_1,c_0} + \lambda_{b_2,c_0}) + \exp(\lambda_{a_0,c_1} + \lambda_{b_1,c_1} + \lambda_{b_2,c_1}) = \exp(0.01) + \exp(0.202)$$

Then:

$$p(c_0|x) = \frac{\exp(\lambda_{a_0,c_0} + \lambda_{b_1,c_0} + \lambda_{b_2,c_0})}{Z(x)} = \frac{\exp(0.01)}{\exp(0.01) + \exp(0.202)} \approx 0.452$$

$$p(c_1|x) = \frac{\exp(\lambda_{a_0,c_1} + \lambda_{b_1,c_1} + \lambda_{b_2,c_1})}{Z(x)} = \frac{\exp(0.202)}{\exp(0.01) + \exp(0.202)} \approx 0.548$$

Then, $c(x) = c_1$.