Introduction to Information Retrieval and Text Mining Assignment 4

Roman Klinger

Institute for Natural Language Processing, University of Stuttgart

2018-01-18

Task 1 (Naïve Bayes)

Train a Naïve Bayes given the following documents annotated with classes c_1 and c_2. Use Add-One-Smoothing. Provide all parameters for a full model specification.

- c_1 "new year"
- c_1 "holiday year"
- c_2 "work again"
- c_2 "never work"

Given the document

- "never holiday"

Which class is assigned by the model?

Task 1 Solution

c_1 "new year"

c_1 "holiday year"

c_2 "work again"

c_2 "never work"

- Prior: $p(c_1) = p(c_2) = \frac{1}{2}$
- $p(\text{new}|c_1) = p(\text{holiday}|c_1) = \frac{1+1}{4+6}$; $p(\text{year}|c_1) = \frac{2+1}{4+6}$; $p(\text{work}|c_1) = p(\text{never}|c_1) = p(\text{again}|c_1) = \frac{0+1}{4+6}$
- $p(\text{work}|c_2) = \frac{2+1}{4+6}$; $p(\text{again}|c_2) = p(\text{never}|c_2) = \frac{1+1}{4+6}$; $p(\text{new}|c_2) = p(\text{holiday}|c_2) = p(\text{year}|c_2) = \frac{0+1}{4+6}$
- Classification: $d =$ "never holiday"
  - $p(c_1|d) = \frac{1}{2}\frac{1}{10}\frac{2}{10} = \frac{1}{100}$
  - $p(c_2|d) = \frac{1}{2}\frac{2}{10}\frac{1}{10} = \frac{1}{100}$
  - $\Rightarrow$ Undecidable.

Task 3 (Maximum Entropy Classifier)

Given the following features (without making a difference between upper and lower case) and documents:

weight	feature
$\lambda_1 = 0.2$	$f_1(y,x) = 1$ if "$" in x and y = SPAM
$\lambda_2 = -0.1$	$f_2(y,x) = 1$ if "$" in x and y = HAM
$\lambda_3 = 0.5$	$f_3(y,x) = 1$ if "Nigerian" in x and y = SPAM
$\lambda_4 = -0.2$	$f_4(y,x) = 1$ if "Nigerian" in x and y = HAM
$\lambda_5 = -0.1$	$f_5(y,x) = 1$ if "you" in x and y = SPAM
$\lambda_6 = 0.4$	$f_6(y,x) = 1$ if "you" in x and y = HAM
$\lambda_7 = 0.1$	$f_7(y,x) = 1$ if y = SPAM
$\lambda_8 = 0.0$	$f_7(y,x) = 1$ if y = HAM

Class y	document
SPAM	x_1 = $1 million from Nigerian defense minister
SPAM	x_2 = Please contact Nigerian finance minister
SPAM	x_3 = You won $30,000!
SPAM	x_4 = Buy these Ginsu knifes now.
HAM	x_5 = You should send the Nigerian wildlife report.
HAM	x_6 = Thanks for great dinner. I owe you $20.

- What is $p(\text{SPAM}|x_1)$ given the maximum entropy classifier with the specified features and weights?
- Calculate the partial derivative of the log-likelihood of all documents with respect to λ_6!

Task 3 (Maximum Entropy Classifier)

weight	feature
$\lambda_1 = 0.2$	$f_1(y,x) = 1$ if "$" in x and y = SPAM
$\lambda_2 = -0.1$	$f_2(y,x) = 1$ if "$" in x and y = HAM
$\lambda_3 = 0.5$	$f_3(y,x) = 1$ if "Nigerian" in x and y = SPAM
$\lambda_4 = -0.2$	$f_4(y,x) = 1$ if "Nigerian" in x and y = HAM
$\lambda_5 = -0.1$	$f_5(y,x) = 1$ if "you" in x and y = SPAM
$\lambda_6 = 0.4$	$f_6(y,x) = 1$ if "you" in x and y = HAM
$\lambda_7 = 0.1$	$f_7(y,x) = 1$ if y = SPAM
$\lambda_8 = 0.0$	$f_7(y,x) = 1$ if y = HAM

Class y	document
SPAM	x_1 = $1 million from Nigerian defense minister
SPAM	x_2 = Please contact Nigerian finance minister
SPAM	x_3 = You won $30,000!
SPAM	x_4 = Buy these Ginsu knifes now.
HAM	x_5 = You should send the Nigerian wildlife report.
HAM	x_6 = Thanks for great dinner. I owe you $20.

- Calculate $p(\text{SPAM}|x_1)$
- Features hold: f_1, f_3, f_7
- $p(\text{SPAM}|x_1) = \frac{e^{\lambda_1+\lambda_3+\lambda_7}}{e^{\lambda_1+\lambda_3+\lambda_7}+e^{\lambda_2+\lambda_4+\lambda_8}} = \frac{e^{0.2+0.5+0.1}}{e^{0.2+0.5+0.1}+e^{-0.1-0.2+0.0}} = \frac{e^{0.8}}{e^{0.8}+e^{-0.3}} \approx 0.75$

Task 3 (Maximum Entropy Classifier)

weight	feature
$\lambda_1 = 0.2$	$f_1(y,x) = 1$ if "$" in x and y = SPAM
$\lambda_2 = -0.1$	$f_2(y,x) = 1$ if "$" in x and y = HAM
$\lambda_3 = 0.5$	$f_3(y,x) = 1$ if "Nigerian" in x and y = SPAM
$\lambda_4 = -0.2$	$f_4(y,x) = 1$ if "Nigerian" in x and y = HAM
$\lambda_5 = -0.1$	$f_5(y,x) = 1$ if "you" in x and y = SPAM
$\lambda_6 = 0.4$	$f_6(y,x) = 1$ if "you" in x and y = HAM
$\lambda_7 = 0.1$	$f_7(y,x) = 1$ if y = SPAM
$\lambda_8 = 0.0$	$f_7(y,x) = 1$ if y = HAM

Class y	document
SPAM	x_1 = $1 million from Nigerian defense minister
SPAM	x_2 = Please contact Nigerian finance minister
SPAM	x_3 = You won $30,000!
SPAM	x_4 = Buy these Ginsu knifes now.
HAM	x_5 = You should send the Nigerian wildlife report.
HAM	x_6 = Thanks for great dinner. I owe you $20.

- $\frac{\partial p_\lambda(Y|X)}{\partial \lambda_6} =$"empirical feature count" $-$"predicted feature count"
- Empirical: $\sum_{(y,\mathbf{x})\in(Y,X)} f_i(y,\mathbf{x}) = 2$
- Predicted:
$$\sum_{(y,\mathbf{x})\in(Y,X)} \sum_{y'} p_\lambda(y' \mid \mathbf{x}) f_i(y',\mathbf{x}) =$$
$$p(\text{HAM}|x_3) + p(\text{HAM}|x_5) + p(\text{HAM}|x_6)$$

Task 3 (Maximum Entropy Classifier)

weight	feature
$\lambda_1 = 0.2$	$f_1(y,x) = 1$ if "$" in x and y = SPAM
$\lambda_2 = -0.1$	$f_2(y,x) = 1$ if "$" in x and y = HAM
$\lambda_3 = 0.5$	$f_3(y,x) = 1$ if "Nigerian" in x and y = SPAM
$\lambda_4 = -0.2$	$f_4(y,x) = 1$ if "Nigerian" in x and y = HAM
$\lambda_5 = -0.1$	$f_5(y,x) = 1$ if "you" in x and y = SPAM
$\lambda_6 = 0.4$	$f_6(y,x) = 1$ if "you" in x and y = HAM
$\lambda_7 = 0.1$	$f_7(y,x) = 1$ if y = SPAM
$\lambda_8 = 0.0$	$f_7(y,x) = 1$ if y = HAM

Class y	document
SPAM	x_1 = $1 million from Nigerian defense minister
SPAM	x_2 = Please contact Nigerian finance minister
SPAM	x_3 = You won $30,000!
SPAM	x_4 = Buy these Ginsu knifes now.
HAM	x_5 = You should send the Nigerian wildlife report.
HAM	x_6 = Thanks for great dinner. I owe you $20.

- $p(\text{HAM}|x_3) + p(\text{HAM}|x_5) + p(\text{HAM}|x_6)$
- $= \frac{e^{-0.1+0.4}}{e^{-0.1+0.4}+e^{0.2-0.1+0.1}} + \frac{e^{-0.2+0.4}}{e^{-0.2+0.4}+e^{0.5-0.1+0.1}} + \frac{e^{-0.1+0.4}}{e^{-0.1+0.4}+e^{0.2-0.1+0.1}}$
- $\approx 0.52 + 0.43 + 0.52 = 1.48$
- "empirical feature count" − "predicted feature count"
  $= 2 - 1.48 = 0.52$

Task 3 (Evaluation)

Explain in your own words what the difference between macro and micro averaging is, when calculating the F measure!
Please make an example with 10 instances and three different classes which shows a lower micro average F score than macro average F score.
It is sufficient to list ten combinations of gold and predicted classes. Explain why your solution leads to the proposed relationship between micro and macro F score.

Task 3 (Evaluation) Solution

- Explain in your own words what the difference between macro and micro averaging is, when calculating the F measure!
  - Both measures aggregate the results by F measures for two or more classes. Micro takes into account the distribution in the data and prediction, macro does not.

Task 3 (Evaluation) Solution

- Please make an example with 10 instances and three different classes which shows a lower micro average F score than macro average F score.

ID	Gold	Prediction
1	A	A
2	B	B
3	C	A
4	C	A
5	C	A
6	C	A
7	C	B
8	C	B
9	C	B
10	C	B

- A:
  - TP=1, FP=4, FN=0
  - Recall=1, Precision=0.2, F=0.4/1.2=0.33
- B:
  - TP=1, FP=4, FN=0
  - Recall=1, Precision=0.2, F=0.4/1.2=0.33
- C:
  - TP=0, FP=0, FN=8
  - Recall=0, Precision=0, F=0
- Macro-average F=0.22
- Micro-average
  - $P = \frac{\mathrm{TP}}{\mathrm{TP+FP}} = \frac{2}{2+4+4} = 0.2$
  - $R = \frac{\mathrm{TP}}{\mathrm{TP+FN}} = \frac{2}{2+8} = 0.2$
  - $F = \frac{2PR}{P+R} = \frac{2 \cdot 0.2 \cdot 0.2}{0.2+0.2} = 0.08/0.4 = 0.2$

Overview

1. Task 1 (Naïve Bayes)
2. Task 2 (Maximum Entropy Classifier)
3. Task 3 (Evaluation)