

Computational Linguistics

Exercise on Maximum Entropy Models

Due: Monday, June 1

2 Generalized Iterative Scaling (GIS)

Show that the distribution in the table below is a fixed point for generalized iterative scaling. That is, computing one iteration should leave the distribution unchanged.

x	y				
profit	„earnings“	f_1	f_2	$\beta = f_1 \log_2(\alpha_1) + f_2 \log_2(\alpha_2)$	2^β
0	0	0	1	1	2
0	1	0	1	1	2
1	0	0	1	1	2
1	1	1	0	2	4

This is an example of a maximum entropy distribution. The corresponding article is from a text classification task. x consists of a single element, indicating the presence or absence of the word *profit* in the article. There are two possible values for y (the article is a member (class) of „earnings“ or not). Feature f_1 is 1 if and only if the article is in „earnings“ and *profit* occurs. Feature f_2 is one in all other cases. We choose as initial parameters $\log_2(\alpha_1) = 2$ and $\log_2(\alpha_2) = 1$. An example of a data set with the same empirical distribution is $((0,0), (0,1), (1,0), (1,1), (1,1))$. Please notice that you have to take the logarithm to the basis of 2!

[Notation remark: $\lambda_i = \log_2(\alpha_i)$.

For further remarks see Manning/Schütze: „Foundations of Statistical Natural Language Processing“, chapter 16.2, p. 592f]

Send your answers to smomtazi@lsv.uni-saarland.de titled “GIS-exercise”.