

Naive Bayes
Lab Session 1
Machine Learning: Pattern Recognition
Master Artificial Intelligence

Camiel Verschoor	Steven Laan
StudentID: 10017321	StudentID: 6036031
UvAnetID: 6229298	UvAnetID: 6036031
Verschoor@uva.nl	S.Laan@uva.nl

September 25, 2012

1 Feature Selection

1.1 Smoothing

1.1.1 Add- λ smoothing

1.1.2 Good-Turing discounting (?)

1.2 Selection criterion

1.2.1 Linear

$$f(x, y) = |x - y|$$

1.2.2 Zipf

When dealing with natural language, the law of Zipf has to be taken into account. This law or probability distribution shows that there are only a couple of words that have a high frequency. Moreover, the most frequent word has a frequency nearly twice as high as the second most frequent. On the other hand, there is massive amount of words that have only been observed once. The zipf distribution then tells us that there is an even bigger amount of words that haven't even been observed once. When plotted on a double logarithmic scale, also known as loglog-scale, the Zipf distribution is a straight line. In figure 1 it can be seen that the data used for training follows a Zipf distribution, as it is more or less a straight line when plotted on a double logarithmic scale.

Based on this distribution, it is probably a better idea to do something with this distribution, instead of discounting in linear fashion.

Therefore, we created the following function:

$$f(x, y) = \frac{1}{1 - |x - y|}$$

