Final Project
CSE 517: Natural Language Processing

Qiao Zhang, Haichen Shen, Danyang Zhuo

March 6, 2016

Question 1: Show that $P(\boldsymbol{c} \mid natural)$ is monotonic in $P(natural \mid \boldsymbol{c})$, under reasonable assumptions.

Answer 1: Suppose we have text $\boldsymbol{c}$ and two naturalness scores, $n_1 > n_2$.

Bayes rule tells us:

$$P(\boldsymbol{c} \mid n_1) = \frac{P(n_1 \mid \boldsymbol{c})P(\boldsymbol{c})}{P(n_1)}$$

$$P(\boldsymbol{c} \mid n_2) = \frac{P(n_2 \mid \boldsymbol{c})P(\boldsymbol{c})}{P(n_2)}$$

Now, we get

$$\frac{P(\boldsymbol{c} \mid n_1)}{P(\boldsymbol{c} \mid n_2)} = \frac{P(n_1 \mid \boldsymbol{c})}{P(n_2 \mid \boldsymbol{c})}\frac{P(n_2)}{P(n_1)}$$

If

$$\frac{P(n_1 \mid \boldsymbol{c})}{P(n_2 \mid \boldsymbol{c})} > 1$$

and we make the assumption that the naturalness score is uniformly distributed

$$\frac{P(n_2)}{P(n_1)} = 1$$

then, we get the monotonic property we want,

$$\frac{P(\boldsymbol{c} \mid n_1)}{P(\boldsymbol{c} \mid n_2)} > 1$$

The monotonicity property means that the more natural a text in truth is, the higher data likelihood (the lower perplexity) for the text we would observe.