COMS W4705
Natural Language Processing
Homework 1

Shane Chin

sac2171

1   Question One

$$q(w_i|w_{i-2}, w_{i-1}) = \lambda_1 \times q_{ML}(w_i|w_{i-2}, w_{i-1}) + \lambda_2 \times q_{ML}(w_i|w_{i-1}) + \lambda_3 \times q_{ML}(w_i)$$

$$L(\lambda_1, \lambda_2, \lambda_3) = \sum_{w_1, w_2, w_3} c'(w_1, w_2, w_3) \log q(w_3|w_1, w_2)$$

$$q_{ML}(w|u, v) = c(u, v, w)/c(u, v)$$

$$2^{-l}$$

$$l = \frac{1}{m} \sum_{i=1}^{m} \log P(s_i)$$

then, Perplexity is equal to

$$\sqrt[m]{\sum_{i=1}^{m} \log P(s_i)}$$

After using the chain rule to expand the probabilities of each word. as the inner parts of the sum are equivalent, except Perplexity is raised to a negative power, we can see that minimizing Perplexity is maximizing the test set probability.

2   Question Two

The problem in this example is that when you have a composition of sentences the trigram will sometimes be severely underweighted. For example, if we see "The cat jumped" < 6 times, and "The cat meowed" only once, the weight of the trigram estimate will be severely lowered.

3   Question Three

For each word we create our matrix of most likely path probabilities however the partial best path is told to us by the trigram, so rather than using the whole estimate, only use the trigram estimate. Instead of State partial probabilities you have trigram possibilities at each tick.