

Weekly Homework 5

Dalton Rothenberger

- A
 - a) - This model does not accurately translate phrases longer than a single word. One example is when the author talks about "will be" and "seront". The author feels that "will" and "be" act in combination to produce "seront" but the model aligned "will" with "seront" and "be" with nothing. Also, in the section 7 it is shown that only 30% of the results were exact or alternate meanings where as the rest were invalid translations which shows this model is not very accurate.
 - b) - The Markov assumption is being used when translating. The translator does not look at the surrounding n-grams for info, it only looks at the current n-gram for information to use in the translation.
 - c) - They state that the bigram model has a perplexity of 78 and the perplexity of the trigram model is 9. This means the trigram model will predict the sample better than the bigram model did.
 - d) - The model uses fertility in combination with distortions to determine whether an English word translates to multiple French words and what order in the sentence should these words appear. This relates to the Noisy Channel model because the model has to look at the French sentence and unscramble it so that the grammar makes sense in English. For example, it has to unscramble the adjectives coming before or after nouns depending on language.
 - e) - The prior signifies the order in which to place source words. The posterior signifies words from the source language that might have produced the observed words.
 - f) - $\hat{S} = \underset{s}{\operatorname{argmax}} Pr(S|T)$
 - g) - The parameters of the translation model are a set of fertility probabilities for each English word and for each fertility a moderate limit (25 in this paper); a set of translation probabilities, one for each element of French vocab and each member of the English vocab; and a set of distortion probabilities.
 - h) - A distortion is when a word in one language appears in a different spot than the word that produced it in the other language. Example: Adjectives coming before or after nouns.
 - i) - An alignment indicates the origin in the original sentence for the words in the translated sentence.
- B
 - a) - Graphically, the bias term serves as the shift on the logistic regression. In 2D it serves as the b value in $y = mx + b$. In 3D moves the plane around.
 - b) In gradient descent the whole training set is considered before taking one model parameters update step. In stochastic gradient descent only one data point is considered for each model parameters update step, cycling over the training set. SGD is at least as good as gradient descent
 - c) - The gradient
 - d) - Features that were not present in the current iteration/data entry.
 - e) - The partial derivatives with respect to each of the independent variables. The direction to move in to go towards the optimal solution for each of those dimensions.