NLP Quiz 2 solution_ Rubric

Question 1:

Binary

$P(x_i|x_{i-1},x_{i-2},...,x_1)=P(x_i|x_{i-1},x_{i-2})$

4 Order Markov chain will **capture more context t**han 2nd order Markov Chain if corpus size is large enough.

.

Question 2:

nGram model works better if test corpus is very much similar to the training corpus, otherwise many nGram will be missing in training corpus making less efficient prediction.(1 marks) .
For reasoning which is little correct (½ marks)

Solution: smoothing , interploation , back-off (1 marks)

Question 3: intuition for perplexity:

It is the weighted branching factor or can be thought of as number of choices one have after a particular decision. (1)

When training data is not like testing data, it do not give true essence(1)

Question 4:

Binary

Ans: ((((3/10)^3)*((2/10)^2)*((1/10)^5))^-1/10)

Binary(2)

Question 5: This question is based on simple interpolation concept

Take the estimate $q(w_i|w_{i-2},w_{i-1})$ to be:

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

P(Dr| Saikat Singh)=>.5 * p(Dr| Saikat Singh) + .3* p(Dr | Saikat) + .2* P(Dr)

P(Prof| Saikat Singh)=>.5 * p(Prof | Saikat Singh) + .3* p(Prof | Saikat) + .2* P(Prof)

Upto this .5

5*(1/1) + .3* (1/2) + .2 *(2/60) DR

5*(0) + .3* (1/2) + .2 *(3/60) Prof

Correct solution upto this 2 marks

Larger the probability is the most probable profession of Saikat Singh (Dr)

Question 6: Binary

Total number of cases = 52C3

One card each should be selected from a different suit. The three suits can be chosen in 4C3 was

The cards can be selected in a total of (4C3) x (13C1) x (13C1) x (13C1)

Probability = 4C3 x (13C1)3 / 52C3

= 4 x (13)3 / 52C3

With replacement [0.3976]

With replacement = (3/8)