CS698O: Quiz-1



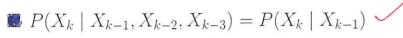


Name: Sohil Dhull

Roll No.: 160607

Please fill the square with blue ink. Check/Tick marks are NOT allowed. Each question is of 2 marks.

1. Consider sequence of R.V.s in the following order: $X_1 \to X_2, \ldots, \to X_n$. Assume that these RV.s follow first order Markov assumption. Which of the following are correct? Fill all that you think are correct.





 $P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i \mid X_{i-1})$



 $\square P(X_1, X_2, X_3) = P(X_3 \mid X_2) \times P(X_2, X_1)$

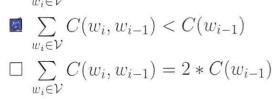
 $\Box P(X_3 \mid X_2) = P(X_2 \mid X_1)$

2. For a bigram language model, which of the following are true? Fill all that you think are correct.

$$\square \sum_{w_i \in \mathcal{V}} C(w_i, w_{i-1}) = C(w_{i-1})$$

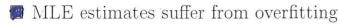
$$\square \sum_{w_i \in \mathcal{V}} C(w_i, w_{i-1}) > C(w_{i-1})$$

$$\sum_{w_i \in \mathcal{V}} C(w_i, w_{i-1}) < C(w_{i-1})$$

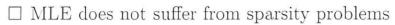


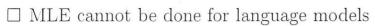


3. Which of the following statements about Maximum Likelihood Estimate (MLE) for Language Model are correct? Fill all that you think are correct.



☐ MLE underestimates probabilities for n-grams with high counts and overestimates probabilities for n-grams with low counts







4. Which of the following are correct MLE estimates for parameters of n(=1 or 2 or 3 or 4)-gram language models? Fill all that you think are correct.

$$\theta_{MLE}(w_i) = \frac{C(w_i)}{|\mathcal{V}|}$$

$$\square \ \theta_{MLE}(w_i \mid w_{i-1}) = \frac{C(w_i, w_{i-1})}{C(w_i)}$$

$$\square \ \theta_{MLE}(w_i \mid w_{i-1}, w_{i-2}) = \frac{C(w_i, w_{i-1}, w_{i-2})}{\sum\limits_{w_i \in \mathcal{V}} C(w_i, w_{i-1}, w_{i-2})}$$





- 5. Which of the following are correct about perplexity measure? Fill all that you think are correct.
 - \square The higher the perplexity of a language model the better
 - The lower the perplexity of a language model the better



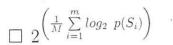




6. Which of the following are correct definition(s) of perplexity? Fill all that you think are correct.

$$2^{-\left(\frac{1}{M}\sum_{i=1}^{m}log_2\ p(S_i)\right)}$$

$$\square \left(\frac{1}{M} \prod_{i=1}^{m} p(S_i)\right)^{-\frac{1}{M}}$$



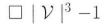


- 7. Which of the following strategies can be used for overcoming MLE limitations for LM? Fill all that you think are correct.
 - Discounting



- ☐ Extrapolation
- Class Based Clustering of Words
- ☐ Look Ahead Technique
- 8. For a trigram language model, what are the **exact** number of parameters that need to be estimated? Assume $\mathcal{V} = \{STOP, \mathcal{V}\}$. Fill all that you think are correct.







 $\square \mid \mathcal{V} \mid$



Logarithm of Word frequenc		thm of wor	d rank in	natural la	
_		thm of wor	d rank in	natural la	
re correct.	tollowing rela	tionships?			
Power law relationship					
☐ Exponential relationship	0)			
☐ Square relationship					
☐ Linear relationship					
Consider the following corpus	S:				
am John				160	
am out today					
ohn I am					
*					
ohn and cat ran					
The cat ran after the mouse				C	
			$o(John \mid S)$	TART, ST	^{7}AR
$\square \frac{0}{2}, \frac{0}{6}$					
	(2))			
$\frac{2}{3}, \frac{2}{7}$				3	
$\square \frac{0}{2}, \frac{1}{7}$		*			
\ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \	☐ Exponential relationship ☐ Square relationship ☐ Linear relationship ☐ onsider the following corpus am John am out today ohn I am Iary I am he cat ran ohn and cat ran he cat ran he cat ran after the mouse That is the MLE estimate of p	Exponential relationship Square relationship Linear relationship onsider the following corpus: am John am out today ohn I am fary I am he cat ran ohn and cat ran he cat ran after the mouse That is the MLE estimate of $p(STOP \mid cat)$ espectively. Fill all that you think are corpus $\frac{0}{3}, \frac{0}{6}$ $\frac{2}{4}, \frac{2}{7}$	Exponential relationship Square relationship Linear relationship onsider the following corpus: am John am out today ohn I am fary I am he cat ran ohn and cat ran he cat ran after the mouse What is the MLE estimate of $p(STOP \mid cat, ran)$ and $p(STOP \mid cat, ran$	Exponential relationship Square relationship Linear relationship onsider the following corpus: am John am out today ohn I am fary I am he cat ran ohn and cat ran he cat ran after the mouse What is the MLE estimate of $p(STOP \mid cat, ran)$ and $p(John \mid S)$ espectively. Fill all that you think are correct. $\frac{0}{3}, \frac{0}{6}$ $\frac{2}{4}, \frac{2}{7}$	□ Exponential relationship □ Square relationship □ Linear relationship □ onsider the following corpus: am John am out today ohn I am lary I am he cat ran ohn and cat ran he cat ran after the mouse What is the MLE estimate of $p(STOP \mid cat, ran)$ and $p(John \mid START, ST)$ espectively. Fill all that you think are correct. □ $\frac{0}{3}$, $\frac{0}{6}$ □ $\frac{2}{4}$, $\frac{2}{7}$