# **CS 335:** Introduction to Large Language Models

*Habib University*

## Activity Sheet 02

Name: _____ ID: _____

**Question 01: Temperature**

let's consider a simple example with four possible outcomes: 'cheese', 'room', 'it', and 'in'. Let's say we have a probability distribution

$$p(x_i \mid x_{1:i-1})$$

Over these four outcomes such that

$$\begin{cases} p(\text{"cheese"} \mid x_{1:i-1}) = 0.1 \\ p(\text{"room"} \mid x_{1:i-1}) = 0.2 \\ p(\text{"it"} \mid x_{1:i-1}) = 0.3 \\ p(\text{"in"} \mid x_{1:i-1}) = 0.4 \end{cases}$$

Now Consider

$$p(x_i \mid x_{1:i-1})^{1/T}$$

**(a)** What can we conclude about the distribution as $T$ approaches 1 ($T \to 1$).

**(b)** What can we conclude about the distribution as $T$ approaches 0 ($T \to 0$).

**(c)** What can we conclude about the distribution as $T$ approaches $\infty$ ($T \to \infty$).

(You may have to normalize the distribution)

**Question 02: Perplexity**

Perplexity can be interpreted as the average "branching factor" per token. In other words, Perplexity is a measure of how "surprised" the model is when predicting the next word.

$$PP(x_{1:L}) = \exp\left(\frac{1}{L}\sum_{i=1}^{L}\log\left(\frac{1}{p(x_i|x_{1:i-1})}\right)\right)$$

We have a language model that assigns probabilities

$$\begin{cases} p(\text{the}) = \alpha_1 \\ p(\text{cat} \mid \text{the}) = \alpha_2 \\ p(\text{sat} \mid \text{the cat}) = \alpha_3 \end{cases}$$

(a) Calculate the perplexity of test sequence "the cat sat" when $\alpha_1 = 0.4$, $\alpha_2 = 0.6$, $\alpha_3 = 0.8$.

(b) Calculate the perplexity of test sequence “the cat sat” when $\alpha_1 = 0.3, \alpha_2 = 0.4, \alpha_3 = 0.6$.

(c) Calculate the perplexity of test sequence “the cat sat” when $\alpha_1 = \alpha_2 = \alpha_3 = \frac{1}{3}$

(d) When the perplexity is higher, is the language model “more sure” or “less sure” about which word to choose? Explain.