AMS691.02: Natural Language Processing – Fall 2024
Assignment 3

Vishesh Kumar
SBU ID: 115972814

Problem 2.1: Softmax (15 points)

a) Calculate the value of p_k for each $k = 1, 2, 3$

Given $n = 3$ and $s_k = 2 \log k$ for $k = 1, 2, 3$:

Compute $\exp(s_k)$:

$$\begin{aligned}
\exp(s_1) &= \exp(2 \log 1) = \exp(0) = 1 \\
\exp(s_2) &= \exp(2 \log 2) = \exp(2 \times \ln 2) = \exp(\ln 4) = 4 \\
\exp(s_3) &= \exp(2 \log 3) = \exp(2 \times \ln 3) = \exp(\ln 9) = 9
\end{aligned}$$

Sum the exponentials:

$$\text{Total sum} = \exp(s_1) + \exp(s_2) + \exp(s_3) = 1 + 4 + 9 = 14$$

Calculate $p_k = \dfrac{\exp(s_k)}{\sum_{j=1}^{3} \exp(s_j)}$:

$$\begin{aligned}
p_1 &= \frac{1}{14} \\
p_2 &= \frac{4}{14} = \frac{2}{7} \\
p_3 &= \frac{9}{14}
\end{aligned}$$

b) Derive p_n in terms of p_{n-1}, s_{n-1}, and s_n

Let $S_{n-1} = \sum_{j=1}^{n-1} \exp(s_j)$. Then:

$$p_{n-1} = \frac{\exp(s_{n-1})}{S_{n-1}} \implies S_{n-1} = \frac{\exp(s_{n-1})}{p_{n-1}}$$

The total sum Z for n terms is:

$$Z = S_{n-1} + \exp(s_n) = \frac{\exp(s_{n-1})}{p_{n-1}} + \exp(s_n)$$

Compute p_n:

$$p_n = \frac{\exp(s_n)}{Z} = \frac{\exp(s_n)}{\frac{\exp(s_{n-1})}{p_{n-1}} + \exp(s_n)}$$

Simplify:

$$p_n = \frac{1}{1 + \frac{\exp(s_{n-1} - s_n)}{p_{n-1}}}$$

c) Effect of temperature τ as $\tau \to 0$

Adjusted softmax:

$$p_i = \frac{\exp(s_i/\tau)}{\sum_{j=1}^{n} \exp(s_j/\tau)}$$

As $\tau \to 0$:

- For the largest s_i, $\frac{s_i}{\tau} \to \infty$, so $\exp(s_i/\tau) \to \infty$. - For smaller s_i, $\frac{s_i}{\tau} \to -\infty$, so $\exp(s_i/\tau) \to 0$.

Thus, the largest s_i dominates the sum, and p_i approaches 1 for the largest s_i, and 0 for others.

Conclusion: The distribution becomes **sharper** as $\tau \to 0$, concentrating probability mass on the highest s_i.

d) Show that $\frac{\partial p_i}{\partial s_i} = p_i(1 - p_i)$

We have:

$$p_i = \frac{\exp(s_i)}{Z}, \quad \text{where } Z = \sum_{k=1}^{n} \exp(s_k)$$

Differentiate p_i with respect to s_i:

$$\frac{\partial p_i}{\partial s_i} = \frac{\exp(s_i) \cdot Z - \exp(s_i) \cdot \exp(s_i)}{Z^2}$$
$$= \frac{\exp(s_i)[Z - \exp(s_i)]}{Z^2}$$

Since $\exp(s_i) = p_i Z$, and $Z - \exp(s_i) = Z(1 - p_i)$, substitute back:

$$\frac{\partial p_i}{\partial s_i} = \frac{p_i Z \cdot Z(1 - p_i)}{Z^2}$$
$$= p_i(1 - p_i)$$

Problem 2.2: Perplexity (15 points)

The perplexity of a language model on some held-out data X is defined as:

$$\text{ppl} = 2^{-\ell}, \quad \text{where } \ell = \frac{1}{\sum_{x \in X} |x|} \sum_{x \in X} \log_2 p(x)$$

Here, x denotes a sentence, and $|x|$ denotes the number of tokens in x.

a) Compute ℓ and the perplexity

Given:

- Two sentences in X, each with 4 tokens. - Each sentence has probability $p(x) = 0.25$.

Total number of tokens:

$$N = |x_1| + |x_2| = 4 + 4 = 8$$

Sum of log probabilities:

$$\sum_{x \in X} \log_2 p(x) = \log_2 0.25 + \log_2 0.25 = -2 + (-2) = -4$$

Compute ℓ:

$$\ell = \frac{1}{N} \sum_{x \in X} \log_2 p(x) = \frac{-4}{8} = -\frac{1}{2}$$

Compute perplexity:

$$\text{ppl} = 2^{-\ell} = 2^{1/2} = \sqrt{2} \approx 1.4142$$

b) Perplexity of a uniform language model

In this language model, the probability of each token is $p(w_i) = \frac{1}{|V|}$, where $|V|$ is the vocabulary size.

Compute ℓ:

$$\ell = \frac{1}{N}\sum_{i=1}^{N}\log_2 p(w_i) = \frac{1}{N}\left(N \cdot \log_2 \frac{1}{|V|}\right) = -\log_2 |V|$$

Compute perplexity:

$$\text{ppl} = 2^{-\ell} = 2^{\log_2 |V|} = |V|$$

Answer: The perplexity of this language model on X is equal to the vocabulary size $|V|$.