Artificial Intelligence

Exercises week 9 Solution- NLP

Maryam Hashemi

COMP3411/9814

Question 1: Regular Expression

When you create a password on a website often there are constraints on the nature of the password. Suppose the constraints on the password are:

1. Password should be at least 8 characters long but not more than 15 characters.
2. It must have at least one special character from !\%@*.
3. It must have some letters and some digits - that is at least one letter and one digit.
4. At least one of the letters should be a capital letter.
5. It should not have characters not listed above.

Assume you had a function called *match(regexp, str)* that returns True or False depending on whether the regular expression *regexp* matches the string *str* or not.
Write an expression using match or any other regular expression language and boolean operations like OR, AND, NOT that will return True if the password satisfies the given constraints and False otherwise.
Hint: egrep has defined character classes: [[:alnum:]] - all alpha-numeric characters, [[:digit:]] - all digits, [[:upper:]] - all upper case characters. Use the character classes to express your answer concisely.

Answer:

The easiest way is to write regular expressions for each clause in the constraints and AND them.

1. match(([[:alnum:]]—[!@\%*])8,15, password) - True if the password contains only alphanumeric characters or the given special characters and is between 8 to 15 characters long.
2. match([[:upper:]], password) - True if password contains at least one upper case letter.
3. match([[:digit:]], password) - True if password contains at least one digit.
4. match([!@\%*], password) - True if password contains at least one special character.
5. NOT match([\ˆA-Z a-z 0-9 !@\%*], password) - True if password does not contain any character other than those mentioned.

An AND of the above five boolean expressions will check the constraints for the password.

Question 2: Multi-head Attention

Consider the sentence X={"The","cat","sat"}. Track the attention of the first token ("The") on all tokens in the input sequence considering attention dimension (d_k) is 2. The word embedding and initial weights are given to you as follow:

$$X = \begin{bmatrix} x_{the} \\ x_{cat} \\ x_{sat} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 2 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}$$

$$W_Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, W_K = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}, W_V = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$

Answer:
Step 1: Calculate $Q = XW_Q$

$$Q = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 2 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 1 & 2 \\ 1 & 2 \end{bmatrix}$$

Query for "The": $q_{the} = (2, 0)$
Query for "cat": $q_{cat} = (1, 2)$
Query for "sat": $q_{sat} = (1, 2)$
Step 2: Calculate $K = XW_K$

$$K = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 2 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & 3 \\ 3 & 1 \end{bmatrix}$$

Step 3: Calculate $V = XW_V$

$$V = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 2 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 2 \end{bmatrix}$$

Step 4: Calculate Scaled score

$$QK^T = \begin{bmatrix} 2 & 0 \\ 1 & 2 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 0 & 3 \\ 1 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 6 \\ 4 & 6 & 5 \\ 4 & 6 & 5 \end{bmatrix}$$

$$Scaled\_score = QK^T/\sqrt{2} = 0.707 \times \begin{bmatrix} 4 & 0 & 6 \\ 4 & 6 & 5 \\ 4 & 6 & 5 \end{bmatrix} = \begin{bmatrix} 2.83 & 0.00 & 4.24 \\ 2.83 & 4.24 & 3.54 \\ 2.83 & 4.24 & 3.54 \end{bmatrix}$$

The attention of "The" to all tokens: $s_{the} \approx (2.83, 0.00, 4.24)$
Step 5: Apply Softmax

We apply the softmax function to the scaled scores of the first row to get the attention weights (a_{the}).

$$Softmax(z_i) = \frac{e^{z_i}}{\Sigma_j e^{z_j}}$$

Using the unrounded scores: $s_{the} \approx (2.83, 0.00, 4.24)$.
$e^{2.828} \approx 16.92$, $e^0 = 1.00$, $e^{4.243} \approx 69.61$. Sum of Exponential: $16.92 + 1.00 + 69.61 \approx 87.53$

$a_{the} = (16.92/87.53, 1.00/87.53, 69.61/87.53) \approx (0.193, 0.011, 0.796)$
The attention weights for the entire sequence are:

$$Attention = Softmax(QK^T/\sqrt{2}) \approx \begin{bmatrix} \mathbf{0.193} & \mathbf{0.011} & \mathbf{0.796} \\ 0.106 & 0.536 & 0.358 \\ 0.106 & 0.536 & 0.358 \end{bmatrix}$$

Step 6: Compute the Context Vector (Z)

The final output is the weighted sum of the Value vectors V, using the Attention weights. $Z = Attention(Q, K, V) = AV$

$$z_{the} = a_{the}V \approx \begin{bmatrix} 0.193 & 0.011 & 0.796 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 2.603 & 1.989 \end{bmatrix}$$

Interpretation: The calculated vector z_{the} is the new, **context-aware representation** for the word "The." It is the actual output of the attention head for that token. It's the original meaning of "The" (v_{The}), but enriched with information drawn from other words in the sentence ("cat" and "sat"). Since the weight for v_{sat} was the highest (0.796), the resulting vector z_{the} has been heavily influenced by the information contained in the "sat" token. So, the new context vector z_{the} is a blend of all the original Value vectors (v_{the}, v_{cat}, v_{sat}), heavily weighted towards v_{sat} (weight = 0.796). This means "The" is paying the most attention to "sat" to derive its new, context-aware representation.