**NLP – assignment 5**

Koren Abitbul - 318796448, Pan Eyal - 208722058

שאלה 1

1.1

המרכיבים השונים הם השאילתה (Q), המפתח (K), ערך (V).

מנגנון הצומי העצמי מחשב את הscore בין השאילתה (Q) לכל מפתח (K).

על ידי: כאשר: .

ציון זה מציין עד כמה כל מילה אחרת (Key) רלוונטית למילה הנוכחית (Query).

כך, בהינתן קשת התלות הרלוונטית , נסתכל פשוט על הscore: .

1.2

ייתכן שנמצא קשרי צומי חזקים יותר בשכבות שונות של הרובוטריק.

תכונה אפשרית ראשונה: מרחק התלות.  
המודל עשוי ללכוד תלויות של מילים קרובות יותר זו לזו במשפט כבר בשכבות המוקדמות יותר של הרובוטריק, שם הוא מתמקד באינטראקציות מקומיות של מילים. תלויות המילים הרחוקות זו מזו לא יהיו חזקות בשכבות הראשונות כמו בעמוקות יותר.

תכונה אפשרית שניה: מורכבות התלות.  
תלות מורכבת יותר עשויה לחייב את המודל לשלב מידע מחלקים שונים של המשפט. קשרים אלה עשויים להופיע בשכבות מאוחרות יותר, בהן הרובוטריק עיבד מידע דרך שלבי צומי עצמי מרובים ויכול להתמודד עם תלות בטווח ארוך יותר.

1.3

טיוטה:

אופציה אחת היא לעשות קלאסטרים בעזרת פיצרים כמו אורך הקשת של 2 מילים, דירוגי צומי בין שכבות שונות של הרובוטריק עבור שני המילים האלה, בתקווה שכל קלאסטר יכיל קשתות מאותו סוג.

טיוטה לפי gemini: (ועכשיו בשיעור הוא רמז שזה פשוט שיטה שראינו בכיתה)

בעזרת האלגוריתם החמדן לניתוח תלויות, לאחר שנבצע forward pass על הרובוטריק, נשיג את הscore עבור הצומי העצמי בין כל צמדי המילים עבור כל השכבות. נוכל להתחיל עם כל מילה בתור "ראש" של מילה בעץ התלות. לבדוק את הscore עם שאר המילים לפי שכבות שונות ולאט לאט לבנות את התלויות התחביריות.

שאלה 2

2.1

נוכיח שהפונקציה הזאת היא ההופכית של פונקציית הסיגמואיד.

עבור פונקציית הסיגמואיד נסתכל על כאל הפלט של הסיגמואיד בהינתן הקלט ונחפש פונקציה ש הוא הפלט שלה בהינתן :

*2.2*

נעשה את אותו הדבר עבור פונקציית הsoftmax :

קיבלנו תלות ב-, לכן אין לפונקציה הופכי. וזה הגיוני כי לדוגמה אם בsoftmax היינו מקבלים התפלגות אחידה, אז הלוגיטים יכולים להיות כל וקטור אשר הערכים שלו הם זהים.   
בשביל לחשב את הלוגיטים יש צורך בלדעת את הערך:

2.3

בשביל לחשב את הלוגיטים יש צורך בלדעת את הערך:

2.4

אם ניקח את k האסימונים המובילים ישירות מהלוג'יטים, על אף שאותם הטוקנים ייבחרו, ההסתברות הסופית יכולה לצאת מוטה ולא מדוייקת, זאת מכיוון שהלוג'יטים אינם מנורמלים על ידי הsoftmax. ההבדלים ביניהם עשויים להיות גדולים בהרבה מהפרשי ההסתברות בפועל.

לדוגמא: עבור רגיל על softmax כאשר :

ועבוראשר נלקח על הלוגיטים כאשר :

*וקל לראות כי*

3

לא מצאנו ☹