

## שאלה 2

(b) הגענו לציון F1 של 0.8.

## שאלה 3

(b)

על מנת לבצע pruning עשינו בדיקה לכל צמד של מילה ותיוג – אם התיוג אינו מופיע עבור המילה בכלל אז לא ביצענו את שאר החישוב כיוון שתמיד נקבל הסתברות 0.

ראינו שתנאי זה משיג תוצאות טובות (5-6 שניות) ולכן בחרנו במימוש זה ולא מימוש מחמיר יותר (כזה שדורש לפחות  $k > 0$  מופעים של המילה והתיוג יחד).

```
if e_x_v < 1:
    continue
```

בחרנו:  $\lambda_1 = 0.5$ ,  $\lambda_2 = 0.3$ ,  $\lambda_3 = 0.2$ .

(c)

הגענו לציון F1 של 0.84.

(e) נגדיר את קבוצות המילים והתיוגים האפשריים:

$$\mathcal{X} = \{x, y, z\}, \mathcal{Y} = \{A, B\}$$

כעת נבחר שני משפטים ותיוגים עבורם:

$$s_1: x \ y \ y \\ t_1: B \ B \ B$$

$$s_2: x \ z \ x \\ t_2: A \ A \ A$$

והמשפט שאת התיוגים שלו נרצה לחזות:

$$s: x \ y$$

נחשב את ההסתברויות לפי הדוגמאות, נציין רק הסתברויות שאינן אפס (כלומר שראינו לפחות הופעה אחת של הזוג/השלשה המתאימה):

$$\begin{aligned} q(A | *, *) &= \frac{1}{2}, & q(B | *, *) &= \frac{1}{2} \\ q(A | *, A) &= 1 \\ q(A | A, A) &= \frac{1}{2}, & q(STOP | A, A) &= \frac{1}{2} \\ q(B | *, B) &= 1 \\ q(B | B, B) &= \frac{1}{2}, & q(STOP | B, B) &= \frac{1}{2} \end{aligned}$$

$$\begin{aligned} e(x | A) &= \frac{2}{3}, & e(x | B) &= \frac{1}{3} \\ e(z | A) &= \frac{1}{3} \\ e(y | B) &= \frac{2}{3} \end{aligned}$$

כעת נדמה את ריצה האלגוריתם החמדן. בשלב הראשון, האלגוריתם יבחר את התיוג A כיוון ש:

$$e(x | A) \cdot q(A | *, *) = \frac{1}{3} > \frac{1}{6} = e(x | B) \cdot q(B | *, *)$$

בשלב השני נראה שכל בחירה תיתן הסתברות 0:

$$\begin{aligned} e(y \mid A) \cdot q(A \mid *, A) &= 0 \cdot 1 = 0 \\ e(y \mid B) \cdot q(B \mid *, A) &= \frac{2}{3} \cdot 0 = 0 \end{aligned}$$

ולכן נקבל שבכל מקרה ההסתברות של כל התיוג שהאלגוריתם יחזיר תהיה 0. לעומת זאת, אם נבחר את התיוג  $B$ ,  $B$ :

$$p(xy, BB) = q(STOP \mid B, B) \cdot q(B \mid *, *) \cdot e(x \mid B) \cdot q(B \mid *, B) \cdot e(y \mid B) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3} \cdot 1 \cdot \frac{2}{3} \geq 0$$

והוכחנו שהאלגוריתם החמדן לא מחזיר את התיוג בעל ההסתברות המקסימלית.

## שאלה 4

(d) האלגוריתם החמדן שלנו הגיע לציון F1 של 0.73, ואלגוריתם ויטרבי לציון של 0.56.

(e) דגמנו כמה מקרים בהם האלגוריתם החמדן טעה:

```
True tag:
['0', '0', '0', '0', '0', '0', '0', '0', '0', '0', 'PER', 'PER', '0', 'MISC', '0', '0', 'PER', 'PER', '0', '0', '0', '0', '0']
Greedy tag:
['0', '0', '0', '0', '0', '0', '0', '0', '0', '0', 'PER', 'PER', '0', '0', '0', '0', 'PER', 'PER', '0', '0', '0', '0', '0']
True tag:
['LOC', '0', 'ORG', '0', '0', '0', '0', '0', 'ORG', '0', '0', 'PER', 'PER', '0', '0', 'PER', '0', '0', '0', '0', 'PER', 'PER', '0', '0', 'PER', 'PER', '0', '0']
Greedy tag:
['LOC', '0', 'ORG', '0', '0', '0', '0', '0', 'ORG', '0', '0', 'PER', 'PER', '0', '0', 'ORG', '0', '0', '0', '0', '0', '0', 'PER', 'PER', '0', '0', '0', 'ORG', 'ORG', '0']
True tag:
['0', '0', '0', 'MISC', '0', '0']
Greedy tag:
['0', '0', 'PER', 'PER', 'PER', '0']
True tag:
['0', '0', '0', '0', '0', '0', '0', '0', 'ORG', 'ORG', 'ORG', 'ORG', 'ORG']
Greedy tag:
['0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', 'LOC', '0', 'ORG']
```

המסקנה העיקרית בה ניתן להבחין היא שהאלגוריתם נוטה לעשות "טעויות נגררות" – ברגע שבנקודה מסוימת ההסתברות לתיוג לא נכון הייתה גבוהה יותר, ההסתברות שנטעה בתיוגים הבאים גדלה. למשל בדוגמא האחרונה, האלגוריתם לא זיהה שהמילה החמישית מהסוף מתארגת ארגון, ולכן ההסתברות שכל התגיות הבאות יהיו של ארגון פחתה והוא טעה גם בהן. סוג טעויות זה תואם מאוד את אופי האלגוריתם – הוא מקבל החלטות שהן אופטימליות רק לנקודת הזמן הנוכחית, ולכן עלול "להטעות" את עצמן.

## שאלה 5

(a) i. ללא הכפלה בוקטור mask, פונקציית ה-loss כנראה תגדל משמעותית, במיוחד עבור וקטורי קלט קטנים יותר – לפונקציה J כפי שהוגדרה בתחילת השאלה יתווסף הסכום:

$$\sum_{t=T+1}^M CE(y^{(t)}, \hat{y}^{(t)}) = \sum_{t=T+1}^M -\log \widehat{y}_{C+1}^{(t)} > \varepsilon \cdot (M - T - 1)$$

כאשר:

$$\varepsilon := \min_{T+1 \leq t \leq M} -\log \widehat{y}_{C+1}^{(t)}$$

ותחת ההנחה הסבירה שהמודל לא יתן הסתברות 1 לתיג C+1, מתקיים  $\varepsilon > 0$  ולכן פונקציית ה-loss תגדל ביחס הפוך לאורך הקלט. בהתאם, עדכון הגרדיאנטים יהיה בעיקר כדי לתקן את תחזיות המודל על וקטורי הריפוד ולגרום לו לחזות את התיג ה-C+1 בודאות גבוהה. שימוש ב-mask פותר בעיה זו כיוון שהוא מאפס את הסכום שכתבנו לעיל ומוריד אותו מ-J, כל שגם עדכוני הגרדיאנטים לא יושפעו ממנו.