

## למידה עמוקה 2022ב

### הערות לפתרון ממ"ן 42

#### שאלה 1

יש הרבה דרכים לשלוט בחישוב הגרדיאנט.

הכי פשוט: כל עוד לא הגענו לטוקן ה-  $T-t$  בלולאה במתודת ה-`forward` של ה-`RNN`, נבצע את כל החישוב בתוך הקונטקסט של `torch.no_grad()`. כך במעבר אחורה הנגזרת תפועפע רק עד הטוקן בו אנחנו מתחילים להתחיל לעקוב אחרי החישובים, בדיוק כדרוש.

#### שאלה 2

חשוב לשים לב שהקלט לתא הנשנה הוא השיכון של הטוקן בעוד שהמצב החבוי באופן מקובל לא נחשב כקלט אלא מאפיין השייך לתא עצמו. לכן בחישוב `R` יש להשתמש רק בשיכון הטוקן הנוכחי.

כמו כן יש לשים לב שהאקטיבציה המתאימה ל-`R` היא סיגומאיד, שכן אנו מעוניינים בערך בין 0 (שכחה מלאה) ל-1 (זכרון מלא). כמובן שהמעבר ביניהם צריך להיות רציף, שכן אנחנו צריכים לחשב נגזרות דרך אקטיבציה זו.

#### שאלה 3

אחת הדרכים האפשריות לדרבן את הרשת ללמוד מבנה כזה היא לבנות את פונקציית המחיר כך שהיא משווה כל שכבה במקודד לשכבה המתאימה במפרש, לא רק את ההתחלה והסוף.

רעיון דומה ומוכר למקודד עצמי מהסוג הזה נקרא `Flow Model`. במאמר על מודלים גנרטיביים המקושר בפרק על שימושים למקודדים עצמיים תוכלו לקרוא עליהם.

#### שאלה 4

קראו על BERT (למשל בבלוג של ג'אי אלאמאר) כדי לראות כיצד פתרו בעיה דומה מאוד בצורה מוצלחת.