

למידה עמוקה 2022ב

הערות לפתרון ממ"ן 12

שאלה 1

בציור דיאגרמת הבלוקים יש לתת ביטוי מפורש לפיצול המתרחש ולשימוש החוזר באותה שכבה. דרך איור זה ניתן להבין שכאשר מחלקים את הקלט ל-4, השימוש החוזר בפרמטרים הופך אותם (פוטנציאלית) למשמעותיים יותר עבור פונקציית המחיר, ואליהם מגיע הגרדיאנט מארבעת השימושים השונים. זו התשובה לסעיף האחרון.

שאלה 2

א.

- הגדרת השאלה מבקשת לאפס בדיוק חצי מהפיצ'רים והאקראיות היא מי הם אלו שיתאפסו. חלקכם מימשו את השכבה כך שבממוצע חצי מהם יתאפסו, בדומה ל-dropout הרגיל. נסו לחשוב על ההבדל במימוש (אפשר למשל להגריל את מי שמתאפס בעזרת randperm).
- שימו לב שמכיוון שמשתמשים במימד המאפיין לנרמל את הנתונים, אין שום בעיה לעשות זאת גם במצב חיזוי (כי אנו לא יוצרים תלות בין דגימות שונות ב-batch). לכן אין צורך בחישוב ממוצע ושונות למטרה זו.

ה.

יש לשאוף להשוות רשתות דומות ככל האפשר, ולכן כדאי להשתמש בשכבות LayerNorm ו-Dropout המובנות במקום שכבות ה-DropNorm שכתבתם אתם. חלקכם השתמשתם בשכבות BatchNorm. זה לא אסון, אבל פחות דומה ולכן פחות מוצלח עבור ההשוואה.

שאלה 3

שימו לב שכאשר אתם דוגמים batch בדרך הבאה (כנראה עקב באג במדריך):

```
next(iter(dataloader))
```

אתם מייצרים איטרטור חדש מ-dataloader ודוגמים ממנו את ה-batch הראשון. אם ה-dataloader מוגדר לדגום באופן אקראי את הדאטא אז הדבר שקול לדגימת batch אקראי של הנתונים, אך בצורה מאוד איטית (כי צריך בכל פעם ליצור מחדש את האובייקט). אם ה-dataloader לא טוען באקראי, תמיד תקבלו את אותו ה-batch הראשון.

רוב הזמן עדיף להגדיר פעם אחת את האיטרטור:

```
iterator = iter(dataloader)
```

ובכל איטרציית אימון לקחת ממנו את ה-batch הבא בתור:

```
batch = next(iterator)
```