

מטלת מנחה (ממ"ן) 12

הקורס: 22961 למידה עמוקה

חומר הלימוד למטלה: יחידות 2-4

מספר השאלות: 3

סמסטר: 2022ב

משקל המטלה: 5 נקודות

מועד אחרון להגשה: 22.04.2022

קיימות שתי חלופות להגשת מטלות:

- שליחת מטלות באמצעות מערכת המטלות המקוונת באתר הבית של הקורס
 - שליחת מטלות באמצעות הדואר או הגשה ישירה למנחה במפגשי ההנחיה
- הסבר מפורט ב"נוהל הגשת מטלות מנחה"

שאלה 1 (25 נקודות)

בשאלה זו עליכם לממש שכבת רשת ב-PyTorch בשם SplitLinear.

א. ממשו את המעבר קדימה ברשת כמפורט להלן:

- הקלט X הוא batch של דגימות של טנזור חד מימדי (משוטח).
 - השכבה תפצל את הקלט באמצעו לשני טנזורים באורך שווה (ניתן להניח שהקלט באורך זוגי).
 - שני החצאים יוזנו בו זמנית דרך אותה שכבת אגרגציה ליניארית ואחרי כן יועברו לאקטיבציית ReLU.
 - לבסוף החצאים ישורשרו בסדר המקורי ליצירת הפלט Y .
- ב. ממשו את הבנאי של שכבה זו, תוך התחשבות באתחול סביר של פרמטרי הרשת. נמקו שיקולכם.
- ג. ציירו דיאגרמה סכמתית בצורת בלוקים של שכבה זו.
- בהמשך השאלה הניחו כי וקטור הקלט הוא בגודל $N \times M$ כאשר N הוא גודל ה-batch ו- M הוא מספר המאפיינים (features, קואורדינטות) בטנזור הקלט.
- ד. חשבו את מספר הפרמטרים בשכבה זו. השוו מספר זה למספר הפרמטרים של שכבה ליניארית רגילה המתאימה לקלט בעל אותם מימדים.
- ה. חשבו באופן אנליטי את הגרדיאנט של פונקציית המחיר ביחס לפרמטרי השכבה: הניחו ש- $\frac{\partial C}{\partial Y}$ כבר חושב וחשבו בעזרתו את $\frac{\partial C}{\partial W}$, $\frac{\partial C}{\partial b}$ כאשר W הם משקלי השכבה הליניארית ו- b וקטור ה-bias.
- ו. ענו על השאלה הבאה מבלי לבצע חישובים חדשים. לו היינו מפצלים את הקלט ל-4 חלקים שווים במקום שניים, כיצד היה משתנה הגרדיאנט של פונקציית המחיר לפי פרמטרי השכבה?

שאלה 2 (50 נקודות)

בשאלה זו עליכם לממש שכבת רשת ב-PyTorch בשם DropNorm, ולהשתמש בה למשימת חיזוי.
א. ממשו את המעבר קדימה בשכבה **במצב אימון** כמפורט להלן:

- הקלט הוא batch של דגימות (טנזור הקלט יכול להיות בעל כל מימד, אין להניח שהוא דו מימדי בלבד).
- השכבה תבחר באקראי, עבור כל הזנה קדימה ברשת, חצי מהמאפיינים בטנזור הקלט ותאפס אותם. יש לשים לב שמדובר בבחירה זהה עבור כל הדגימות ב-batch.
- את שאר המאפיינים בטנזור הקלט השכבה תנרמל לפי הנוסחה

$$x_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}}$$

כאשר:

- x_i הוא מאפיין כלשהו בטנזור הקלט.
- σ^2, μ הם הממוצע והשונות המחושבים על פני כל המאפיינים שלא אופסו בשלב 2. שימו לב שיש לחשבם לכל דגימה ב-batch בנפרד.
- ϵ הוא קבוע קטן המונע חלוקה באפס.
- לבסוף, פלט השכבה (עבור הקואורדינטות שלא אופסו) יהיה

$$y_i = \gamma_i x_i + \beta_i$$

כאשר γ_i, β_i הם פרמטרים נלמדים של השכבה (לכל מאפיין בטנזור הקלט קיים זוג פרמטרים כאלו).

- תכננו וממשו את המעבר קדימה ברשת **במצב חיזוי**, בהתאם לשיקולים הרלוונטיים לשכבות Dropout ושכבות נורמליזציה המופיעים במדריך הלמידה. נמקו תשובתכם.
- ממשו את הבנאי של שכבה זו, תוך התחשבות באתחול סביר של פרמטרי הרשת. נמקו שיקולכם.
- אמנו רשת עמוקה לחיזוי מחלקות פרטי הלבוש באוסף הנתונים Fashion-MNIST, תוך כדי שימוש בשכבות נורמליזציה ושכבות Dropout המובנות בספרייה PyTorch. בצעו אימון זה על גבי GPU.
- חזרו על סעיף ד' אך כעת במקום השכבות המובנות – השתמשו בשכבה אשר מימשתם בסעיפים א'-ג'.

- השוו את התוצאות שהתקבלו בשני הסעיפים הקודמים. באיזו רשת תעדיפו להשתמש?
- בהנחה שגודל טנזור הקלט הוא $N \times 4$ כאשר המימד הראשון הוא מימד ה-batch ו- N הוא גודל ה-batch, חשבו באופן אנליטי את הפעפוע לאחור של הגרדיאנט דרך השכבה הנ"ל. שימו

לב שעליכם לחשב את $\frac{\partial C}{\partial x_i}$ לכל i .

שאלה 3 (25 נקודות)

בשאלה זו עליכם להגדיר DataLoader לאוסף נתונים טבלאי, ולהשתמש בו לאימון רשת רגרסיה.

א. הורידו את אוסף הנתונים Diabetes מאתר הקורס (מופיע בלשונית ממ"ן 12).

ב. קראו על משמעות המשתנים באוסף הנתונים בקישור הבא :

https://scikit-learn.org/stable/datasets/toy_dataset.html#diabetes-dataset

ג. צרו אובייקט Dataset של PyTorch העוטף אוסף נתונים זה.

תזכורת: עליכם להגדיר מחלקה חדשה, היורשת מהמחלקה Dataset ובתוכה להגדיר שלוש מתודות:

- בנאי.
- מתודה המחזירה איבר מהאוסף.
- מתודה המחזירה את מספר הדגימות באוסף הנתונים.

רמזים:

- מומלץ להשתמש בפונקציה `read_csv()` של הספרייה `pandas` לטעינת אוסף הנתונים לזכרון בעת אתחול האובייקט.
- אם אתם משתמשים ב-Colab, תוכלו להנגיש את הקובץ לסביבת העבודה בענן בעזרת הפקודה `mount(path)` מהספרייה `google.colab.drive` ומיקום הקובץ ב-Google Drive.

הערה: בשאלה זו אסור לכם להשתמש בפונקציות המובנות של `sklearn` לטעינת

אוסף הנתונים – יש לטעון את הנתונים ישירות מקובץ ה-CSV.

ד. צרו אובייקט DataLoader הטוען את הנתונים ב-`minibatches` בגודל 10 דגימות.

ה. הדפיסו `batch` יחיד מאוסף הנתונים.

ו. השתמשו ב-`DataLoader` לאימון רשת נוירונים עמוקה החוזה את המשתנה האחרון באוסף (המשתנה `Y`) על סמך שאר המשתנים.

ז. בדקו את ביצועיה של הרשת על סט בדיקה המורכב מ-20% מאוסף הנתונים (שייבחרו באופן אקראי). שימו לב שאין לאמן את הרשת על נתונים אלו ובהתאם יש לשנות את הגדרת אובייקט ה-Dataset.

יש להגיש בקובץ ZIP אחד :

1. מחברת ג'ופיטר המכילה את הקוד והפלט הדרוש לכל השאלות.
 2. קישור לסרטון קצר בו אתם מסבירים את הקוד, מריצים אותו ומנתחים/מסבירים את הפלט (את הסרטון יש להכין בעזרת תוכנת Zoom ושימוש בפונקציה `Share screen`).
 3. תשובות בכתב ככל שנדרש.
- מומלץ להשתמש במערכת הגשה האינטרנטית.