

קורס "רשתות נוירונים" 83695 – פרויקט סיום

נהלי הגשת הפרויקט:

יש להגיש (ולהכין) את העבודה בזוגות. אין לעבוד בקבוצות גדולות יותר, גם לא לצורך שיתוף רעיונות. האוניברסיטה רואה בחומרה העתקות, כל הגשה שתערב העתקה תטופל בועדת משמעת אוניברסיטאית.

יש לכתוב באופן ברור בתרגיל מי המגשים. על הסטודנטים/יות להגיש את התרגיל דרך הלמדה (כל אחד/ת מבני הזוג). יש להגיש PDF הכולל את הגרפים והתשובות לשאלות. בנוסף יש להגיש את הקוד לשאלות הרלוונטיות, ניתן לכתוב בפייתון או מטלב, מותר להשתמש בספריות מובנות. הגשה בפייתון תהיה של קובץ jupyter הכולל את התוצאות.

יש לתעד את הקוד ולהסביר בקצרה כל סעיף בו אתם נדרשים לממש או לאמן מודל כלשהו. על כל הגרפים לכלול כותרות, סקאלות ויחידות נכונות של הצירים.

תאריך הגשה: 8.9.2024

בהמשך ייקבעו תאריכים להצגת הפרויקט מולנו בזום במהלך ספטמבר.

תיאור הפרויקט:

בפרויקט זה תעבדו עם בסיס נתונים הכולל רישום של אותות חשמליים מנוירונים שונים ותבנו מסווגים לסוגי הנוירונים על-בסיס הפעילות החשמלית שלהם. התאים נאספו ממוחות של בני אדם ועכברים משכבות שונות של קליפת המוח. גירוי חשמלי הוזרם לתוך התאים ביחידות של pA, ותגובה חשמלית של התאים לגירוי נמדדה ביחידות של mV. המידע נאסף ע"י מכון Allen, ניתן לקרוא עוד פרטים על האיסוף כאן <https://celltypes.brain-map.org/overview>.

הפרויקט יכלול מספר חלקים שבהם תבצעו ניתוח של הנתונים על-פי מדדים שונים, ויזואליזציה של תוצאות, ואימון של רשתות ומסווגים שונים.

הקובץ הבא מכיל את בסיס הנתונים עבור החלק הראשון.

הקובץ הבא מכיל את בסיס הנתונים עבור החלקים האחרים.

חלק א

בחלק זה תבצעו היכרות ראשונית עם הנתונים, תנתחו אותם על-פי מדדים שונים, תבצעו ויזואליזציה של התוצאות ותבנו מסווג בסיסי.

תחילה טענו את הקובץ הבא באמצעות הפקודה `np.load` של פייתון. הקובץ מכיל שלושה מערכים אשר תואמים (משמאל לימין) ל גירוי החשמלי ותגובה חשמלית של נוירון מסוג `spiny` ו `aspiny` / נתון כי תדר הדגימה הוא 50 KHz.

1. קראו על נוירונים `spiny` ו-`aspiny` והסבירו בקצרה מהם סוגי נוירונים אלו.
2. הציגו את אות הגירוי החשמלי כתלות בזמן. תארו את הגירוי החשמלי בקצרה.
3. הציגו את הפעילות החשמלית של הנוירון ה-`spiny` והנוירון ה-`aspiny` בתגובה לגירוי החשמלי. תארו בקצרה את התגובה.
4. חשבו והציגו `spike train` של שני הנוירונים בתגובה לגירוי החשמלי. מהו ערך הסף שמעליו מתרחש פוטנציאל פעולה בנוירונים אלו?
5. חשבו והציגו בגרף את ה- `time dependent firing rate` עבור כל אחד מסוגי הנוירונים. השתמשו ב- `bins` וב- `window/kernel` שנראים לכם מתאימים כדי שפונקציית קצב הירי תתאר בצורה טובה את תגובת הנוירונים לגירוי החשמלי. הסבירו את בחירתכם.
6. מהו קצב הירי הממוצע של כל אחד מהנוירונים בתגובה לגירוי החשמלי? הניחו שקצב ירי זה מאפיין את התגובה של כל אחד מסוגי הנוירונים. סביב ממוצע זה, צרו התפלגות גאוסיאנית של ההסתברות לקצבי

ירי בתגובה לגירוי עבור כל אחד מסוגי הנירונים. השתמשו בסטיית תקן לבחירתכם עבור ההתפלגויות כך שיש ביניהן חפיפה חלקית.

(a) הציגו את ההתפלגויות בגרף.

(b) מהו ה-discriminability עבור ההתפלגויות שיצרתם? הראו כיצד חישובתם.

7. בחרו ערך סף z כלשהו שישמש לסיווג בין שני סוגי הנירונים. נתחו את הביצועים של המסווג: הציגו confusion matrix וחשבו מדדים רלוונטיים כגון accuracy, recall וכו'.

8. חשבו והציגו Receiver operating characteristics (ROC) curve עבור הסיווג בין שני סוגי הנירונים על בסיס המסווג הפשוט. מהו ה-Area under the curve שהתקבל ומה משמעותו? איזה שינוי בנתונים שעל בסיסם חישובתם אותו עשוי להביא ל-AUC גדול או קטן יותר?

חלק ב

בחלק זה תשתמשו בשיטות למידה בלתי מפקחת על מנת לעשות אנליזה לנתונים.

השתמשו רק בערכים הנומריים בקובץ הנתונים השני, כלומר התעלמו מהעמודה המכילה את סוג הנירון.

1. ממשו והריצו PCA-Principal Component Analysis עבור הורדת מימד של הנתונים לשלושה מימדים.

(a) הציגו את שלושת הצירים של ה-PCA באמצעות scatter plot והציגו בצבעים שונים את כל אחד מסוגי הנירונים.

(b) מה מייצגים הצירים של PCA?

(c) האם יש צירים המפרידים טוב יותר מאחרים בין סוגי הנירונים?

(d) בכמה צירים של PCA צריך להשתמש כך שניתן יהיה לשחזר את המידע המקורי עד שגיאה של 1% מבחינת Mean squared error?

2. בנו רשת autoencoder עם שלוש שכבות חבויות ופונקציית אקטיבציה מסוג: relu אם מספר הזהות האחרון של בני/ות הזוג הוא זוגי (הכוונה לסכום הספרות האחרונות של שני בני/ות הזוג), ו tanh אחרת. בחרו את מימד ה-bottleneck להיות 3.

(a) אמנו את הרשת על סט האימון, הציגו את גרף השגיאה (עבור לוס של MSE) לאורך האימון.

(b) הציגו את הדאטה בייצוג הנמוך שנלמד ע"י רשת ה-autoencoder. כלומר הציגו scatter plot של שלושת הצירים ב-bottleneck.

(c) האם רשת ה-autoencoder למדה ייצוג טוב יותר להפרדת סוגי התאים לעומת PCA? נמקו או הדגמו.

חלק ג

בחלק זה תאמנו מסווגים כדי להפריד בין נירון spiny ונירון aspiny. השתמשו בקובץ הנתונים מהחלק השני והשתמשו בעמודה המגדירה את סוג הנירון כדי להגדיר את ה-label. כלומר אל תשתמשו בעמודה זו כמאפיין לאימון המסווגים.

חלקו את הדאטה ל train-test באופן אקראי ככה שסט האימון מהווה 80% מהדוגמאות.

1. אמנו מודל רגרסיה לוגיסטית על סט האימון (train) ובחנו אותו על סט הבחינה (test).

(a) נתחו את השגיאות של המודל כלומר הציגו Confusion matrix וחשבו Precision recall כאשר הנירון ה-spiny יוגדרו כמחלקה ה"חיובית".

2. אמנו perceptron על סט האימון ובחנו על סט הבחינה.

(a) השווה בין שני המודלים מי הכליל טוב יותר?

(b) האם המודל סובל מ overfitting?

3. בנו רשת Multi layer perceptron-MLP על מנת לסווג את שני סוגי הנירונים.

4. אמנו את הרשת עם Gradient Descent-GD על סט האימון, הציגו את גרף ה loss עבור האימון ועבור סט ולידציה בגודל 10% מסט האימון.

(a) נתחו את השגיאות של המודל כלומר הציגו Confusion matrix וחשבו Precision recall כאשר הנירון ה-spiny יוגדרו כמחלקה ה"חיובית".

(b) חזרו על הסעיף הקודם אך עם Stochastic Gradient Descent-SGD עם גודל בטש של 10.

- (c) האם המודל סובל מ overfitting ?
- (d) מה ההבדל בין המודלים שאומנו בשני הסעיפים הקודמים? איזה מודל מכליל טוב יותר?

בהצלחה!