

תרגיל קלסיפיקציה של סיגנל EEG

סטודנטים: פייצוויג בראדלי 311183073, קונסטנטין מלצב 336492640, שי פלד 206230526,
ליאור שגיא 204678825, אלון ישראל 308466903.

מרצה: ד"ר דהן ענת

מבוא

בקורס מבוא לחישה ולמידה, למדנו על מגוון נושאים של המוח האנושי וההשראה של המוח האנושי לשיטות שונות של למידה ממוחשבת. מטלה זו הינה תרגיל קלסיפיקציה של סיגנל EEG בין ילדים עם ליקוי ADHD ובלי. אחד הליקויים בילדי ADHD הינו קשב חזותי, לכן הקלטות EEG מתבססות על משימות קשב חזותי. במשימה, הוצגו בפני הילדים תמונות של דמויות מצוירות והם התבקשו לספור אותם. בכל תמונה הוצגו באקראי בין 5 ל-16 דמויות, וגודל הדמויות היה גדול כדי להקל על הנבדקים לראות ולספור את הדמויות. מיד לאחר תגובת הנבדק, הוצגה לו תמונה נוספת, לכן זמן ההקלטה לאורך המשימה היה תלוי במהירות התגובה של הילד. הקלטת EEG בוצעה על ידי 19 ערוצים: Fz, Cz, Pz, C3, T3, C4, T4, Fp1, Fp2, F3, F4, F7, F8, P3, P4, T5, T6, O1, O2. לכל אחד מ-61 הילדים שנבדקו אשר יש להם ADHD ו-60 ילדים ללא, כקבוצת ביקורת. הילדים כולם בגילאים 7-12.

תשובות לשאלות

1. בוצע עיבוד מקדים של הקבצים.

ראשית נבדק מה המספר המינימלי של שורות עבור כל קובץ mat. כדי לבנות מטריצות בגודל קבוע עבור נבדק. בהסבר על הדה סט באתר, מוסבר כי כי לאחר כל תגובה, הראו ישר לילד תמונה, כמו כן רשום שמספר התמונות שכל ילד ראה, תלוי באופן ישר בזמן התגובה שלו, כלומר ילדים עם זמן תגובה שונה ראו מספר תמונות שונה ולכן גם מספר השורות שונה בהתאם בכל הקלטה. על מנת לבנות דטה סט, לקחנו את אותו המספר המינימלי, ולפי מספר זה טענו מטריצות בגודל קבוע כאשר מספר השורות הוא כמספר המינימלי של שורות בין כל ההקלטות ומספר העמודות הוא 19- לפי מספר ה-features (עבור כל אחת מהאלקטרודות Fz, Cz, Pz, C3, T3, C4, T4, Fp1, Fp2, F3, F4, F7, F8, P3, P4, T5, T6, O1, O2).

לאחר שטענו את כל הנתונים מתוך קבצי matn, חישבנו את המידע ההדדי (mutual information) בין כל 19 הפיצ'רים עבור כל ילד. מידע הדדי של שני משתנים מקריים היא ערך המודד את הקשר ביניהם, כלומר מראה עד כמה ידע על תוצאה X מלמדת על תוצאה Y. אם הערך הינו 0 אזי המשתנים בלתי תלויים, כלומר אי אפשר ללמוד כלום ממשתנה אחד לשני, ומנגד אם הערך גבוה זה מצביע על כך שX מלמד לחלוטין על Y. כלומר בסוף תהליך זה, עבור כל ילד קיבלנו מטריצת קורלציה בין כל הסיגנלים. על מנת לבצע את זה, השתמשנו ב `sklearn.metrics.normalized_mutual_info_score`. מפונקציה זאת, קיבלנו מידע הדדי מנורמל, כאשר הנרמול מתבצע על ידי ממוצע חשבוני. בסוף התהליך הזה קיבלנו מטריצה בגודל 19X19. לאחר מכן ביצענו ממוצע עבור כל עמודה במטריצה של כל ילד וכך קיבלנו עבור כל ילד מערך בגודל 19 כאשר כל איבר בו מציג את הממוצע של המידע ההדדי עבור כל אלקטודה. לבסוף בנינו מטריצה בגודל 121 (כמספר הנבדקים: 61 עם ADHD ו-60 בלי). על 19 (כמספר האלקטרודות). ומתוך העמודות בחרנו לקחת את כל האלקטרודות הממוקמות בקדמת המוח- Fz, Fp1, Fp2, F3, F4, F7, F8.

2. מה הם ה FEATURES עבור הקלסיפיקציה?

כמו שהוסבר קודם בנינו מטריצת קורלציה בין כל זוגות הסיגנלים, אבל בחרנו מתוכם את כל הסיגנלים מהאזורים הפרא-מצחיים והמצחיים במוח.

האזור הזה במוח (האונה הפרא פרונטאלי) אחראי לרוב הפעילות של חשיבה, תפיסה, קשב וריכוז, מסיבה זו, והמלצת המרצה ד"ר דהן ענת, בחרנו להתמקד באלקטרודות הממוקמות בקדמת המוח- Fz, Fp1, Fp2, F3, F4, F7, F8.

3. באיזה שיטת קלסיפיקציה השתמשתם (למשל SVM, KEANS וכו)?

בפרויקט שלנו בדקנו 2 שיטות קלסיפיקציה, KNN: k-nearest neighbors -I SVM: Support Vector Machines.

עבור כל אחד מהמודלים הנ"ל ביצענו ניסוי אשר בודק מה הם הפרמטרים המיטביים עבור כל מודל כזה עם הדתה סט שבנינו. בדיקה זאת אפשרית בעזרת GridSearchCV שזאת פונקציה מתוך sklearn.model_selection. פונקציה זאת מבצעת חיפוש ממצה על כל הערכי הפרמטר שציינו. בשאלה הבאה נראה את התוצאות.

4. כיצד בדקתם את התוצאות ומה היו אחוזי ההצלחה?

חילקנו את הדתה סט כך ש-0.2 מהדתה הינו טסט דתה וכל השאר עבור האימון. בדקנו את הציון של כל מודל בעזרת פונקציות שקיימות עבור כל מודל להערכת הציון של המודל, כאשר הפונקציה מקבל את הקלטים מהדתה סט labeli עבור כל קלט.

להלן התוצאות:

SVM

הפרמטרים המיטביים עבור המודל עם הדתה הנוכחי:

Parameters {'C': 1.0, 'gamma': 'scale', 'kernel': 'rbf'}

כאשר:

C: Regularization parameter. The strength of the regularization is inversely proportional to C.

Gamma: Kernel coefficient for 'rbf', 'poly' and 'sigmoid'.

Kernel: Specifies the kernel type to be used in the algorithm.

הציון עבור test dataset: 0.8

Score: 0.8

KNN

הפרמטרים המיטביים עבור המודל עם הדתה הנוכחי:

Parameters {'metric': 'euclidean', 'n_neighbors': 5, 'weights': 'uniform'}

כאשר:

Metric: The distance metric to use for the tree. The default metric is minkowski, and with p=2 is equivalent to the standard Euclidean metric.

n_neighbors: Number of neighbors to use by default for kneighbors queries.

Weights: Weight function used in prediction.

הציון עבור test dataset: 0.64

Score: 0.64

מסקנה: עבור הדתה סט שבנינו, SVM עם הפרמטרים הנתונים נתן לנו את הציון הטוב יותר של 80%.