

"דימות נתונים" יוני 2020 – עבודת גמר

ד"ר רועי יוזביץ

הקדמה:

במסגרת הקורס, מצ"ב פרטים על עבודת הגמר אשר יש להגיש. מטרת העבודה היא להכיר לסטודנט בצורה אישית ובלתי אמצעית את מגוון הנושאים אשר נלמדו בקורס. בין היתר, יש להבין את הנושאים שעברנו עליהם בקורס הקודם בצורה משמעותית וטובה.

משימות על:

במסגרת עבודת הגמר, נעבוד עם שני Datasets. הראשון הינו Fashion Mnist והשני מסופק לכם בנפרד. מטרת העבודה הינה לייצר אלגוריתם לזיהוי הבגד בצורה שיעבוד יפה על ה-TEST SET.

פירוט המשימות:

- נא לעבור על הקורס של git באתר udacity. זו מטלה חובה. בנוסף, יש לשנות את המחברות בגיטהב שלכם דרך commits שונים.
- יש להשתמש בין היתר באלגוריתמים שלמדנו במהלך הסמסטר. בין היתר: PCA, K, ADABOOST, NAÏVE BAYES וכו'.
- מה מידת הדיוק של המודל שלכם יחסית למנצחים בתחרות KAGGLE? במה הוא יותר טוב ממודל 0?
- כמובן שחובה להשתמש בקורס ולידציה.
- שימו לב ש-PCA חשוב מכיוון שמימד הבעייה הינו 784 (28*28). קיימים מאמרים שטוענים שב-MNIST המקורי ניתן לבצע קלסיפיקציה על בסיס פיקסל אחד בלבד (כמו שראינו).
- ניתן ללמוד גם את הנושא T-SNE ולהשתמש בו.
- הראו את המקומות בהם האלגוריתם שלכם טעה.
- נניח שאני רוצה אלגוריתם שלעולם לא טועה בזיהוי מעיל (הוא מזהה את כל המעילים שהוא מקבל כמעילים). איך צריך לשנות את האלגוריתם שלכם בהתאם?
- מי שרוצה יכול בנוסף לנסות ולהריץ מודל של רשת נוירונים אותו הוא למד בעצמו. בכל מקרה, יש להריץ גם את האלגוריתמים המקוריים.

• האם אתם יכולים לייצר עבור ה-

DATA תמונה כזו?

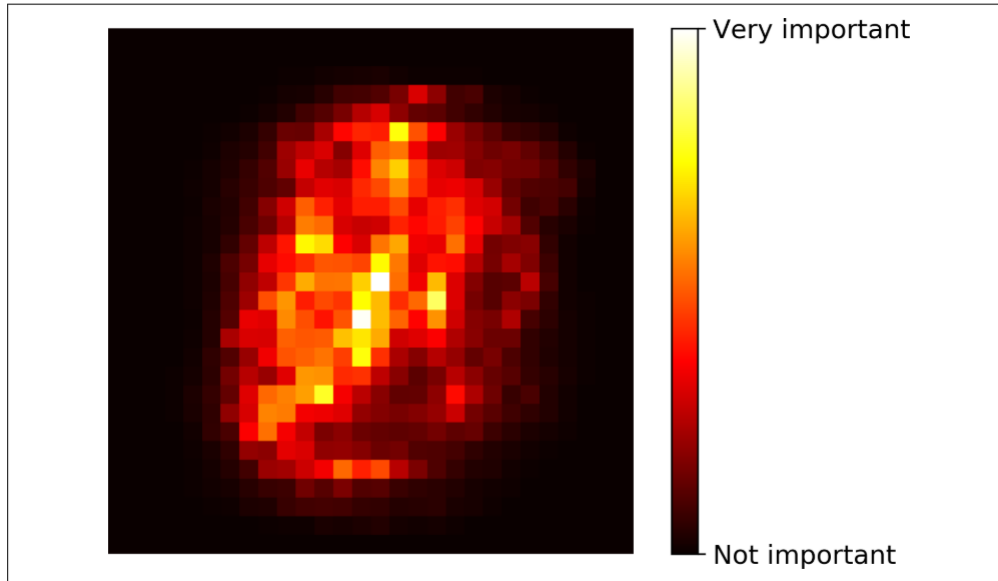


Figure 7-6. MNIST pixel importance (according to a Random Forest classifier)

מקורות לעזרה

- ספר הקורס פרקים 3,4,7,8
- [מאמר על PCA ו-t-SNE](#)
- [מאמר נוסף בנושא](#)

בהצלחה לכולם

מטרה

מטרת העבודה היא לסווג בין שלושה מצבים שונים באופן בו אנשים מתקשרים אחד עם השני. הראשון הינו מצב ספונטני (אוטונומי) בו שני אנשים מזיזים את הידיים שלהם בצורה חופשית אחד מול השני. השני הוא תנועה סינכרונית בה שני האנשים מזיזים את הידיים ביחד והשלישי הוא תנועה במצב לבד. בו רק הצד אחד מזיז את הידיים. הרעיון הוא להסתכל על דפוסי הידיים ולנסות להסיק מהם האם מדובר במצב לבד, ספונטני או סינכרוני.

כפי שראיתם בתרגיל בית מספר 3, יש לכם כבר את כל המידע ב-DATAFRAME מסודר. יש גם להציג את ההבדלים בצורה גרפית יפה. בנוס יינתן ל-3 סטודנטים שיגיעו לרמות הדיוק הגבוהות ביותר. בנוסף, מצ"ב מאמר מדעי שפורסם על העבודה הראשונית הזו. כדאי מאוד לקרוא אותו על מנת להבין מה נצרך מכך בעבודה זו. שימו לב שבתיקה המצורפת, יש תת-תיקיה שנקראת TRAINING ורק עליה יש לעבוד באימון הרשת. בנוסף, יש תת תיקיה הנקראת TESTING ורק עליה יש לעבוד בוולידציה.

הערות נוספות:

- במצב ALONE, קיימת רק הקלטה של יד אחת (יד שמאל). יש לבדוק את המצב הזה אל מול קובץ שנקרא HandRight ובו הקלטה של תנועות ביד ימין. שימו לב שאת כל מצבי ה-ALONE יש לבדוק אל מול הקובץ הזה.
- יש ספריות אשר יש שם 2 מקבצי הקלטות. במקרה כזה, יש לקחת את המקבץ האחרון.
- חשוב מאוד לא לקחת את ה-7 שניות הראשונות של הקובץ.
- הרעיון פה הוא ליצור פרמטר שבודק סינכרון בין שני אנשים. חשבו בעצמכם אילו פרמטרים חשובים ולמה.
- מטרת העבודה היא לסווג תנועות ידיים בין שני אנשים למצב (ALONE, AUTONMOUS ו—סינכרוני). ניתן לחשוב על כל SAMPLE כעל שנייה של הקלטה המורכבת מ-5 timestamps.
- חשוב להבין מה אחוז הדיוק ומהו f-score של האלגוריתם שלכם. ניתן לייצר אלגוריתמים שונים אבל בשום מקרה אסור לאמן את המערכת על המידע מספריית הוולידציה.

בהצלחה לכולם.