

1.

סיווג Naive Bayes מכוסס עם

האוריינטציה הסתברותית. ומשלב בייס נשתק

מפיק עיקרי כאן. מתוך אובייקטים

למקלוליות בק שיתן אסכס עליהם

הם עצמם מהם עליהם. סיווג זה

משתמש ב- words of bag עבור טקסט האובייקטים

מחשב הסתברות של מילה במסגרת בק שנוכח עצמו

כזה, פעמים מילה מופיע בקטל ומה מופך של

באקט מסוים כלומר באיזה קטלוג הוא שייך.

למחצית שני סוגים של מוצגים לסיווג

naive Bayes. כאן זה Multivariate Bernoulli,

ועל זה Multinomial.

במוצג מופל ווריאנטי יש לנו feature אחד

כל מילה במילון שלנו והוא מתקבל ערך בוריאנטי

הקשר אם מילה נמצאת באקט או לא. אולם

זה לא אומר לנו כלום אם מילה אחת נמצאת.

במוצג מופל מוחלט יש לנו feature לכל מ' קיים

המילה באקט. מילים אלה מתוק מילון. אולם

זה לא אומר לנו כלום כי מילים במילון

אחרים. וקיים המילה לא מופיע עם מילון.

מוצג מופל נומינלי בצדק כלל ומציג יותר

יעילה בסיווג לקטלוג. נותן מוצגות יותר

מנוויות למחשבה שמתן ריבה של מופל ווריאנטי

יותר מחוק בגיית האלמנטים והפעלתם. ונבחר

אולם יותר כצדק עצמו פשוט אם מופל

קיים במסגרת או לא.



(2)

1) sklearn, naive\_bayes  
בסיסית זו מצאנו כמה מסוימים של  
naive bayes. נזכר שכל הכי משה'ים לנו.  
Gaussian NB בונה מודל לסיווג עם אלמנטים  
ממדים מסוימים. הוא מקבל את המטריצות הקיימות  
עם סיווג מסוים ומקבל עם המטריצות המדויקות  
כדי למצוא סיווג שביטוי.

Bernall NB אלמנטים בונה מודל לסיווג זיג  
ממדים אחרים וזריק מקבל סיווג בינארי.  
הוא עוזר בשיטה מולטי ווליאני כחומר למחשבה.  
1 - Multinomial NB מודל הכי משה'ים לנו כי הוא  
עוזר לפי שיה משה'ים ווליאני ומקבל ממדים  
לפי vector counts ו- vector counts.

2) KDT

באן מצאנו מודל נקרא Naive Bayes Classifier  
מודל הזה מקבל ממדים אחרים כולם סיווג  
שניהם שהוא העצם בינארי וזה מביא לנו  
שיטה מולטי ווליאני שמצאנו הקבוצה וזה פחות  
משה'ים לנו משה'ים סיווג של ממדים  
שקטוראליים.



NLTK with sklearn wrapper (3)  
 sklearn Classifier - ממשלה  
 ממשלה ספריה  
 nltk.classify.scikitlearn  
 ומכאן אלמנטים  
 ממשלה של sklearn ממשלה  
 ממשלה ממשלה MultinomialNB כמו ממשלה  
 ממשלה ממשלה ממשלה ממשלה ממשלה  
 ממשלה ממשלה ממשלה - sklearn Classifier - ממשלה ממשלה  
 ממשלה ממשלה ממשלה

Textblob.classifiers (4)  
 Naive Bayes Classifier - ממשלה  
 ממשלה ממשלה textblob.classifiers  
 ממשלה ממשלה ממשלה ממשלה ממשלה  
 ממשלה ממשלה ממשלה ממשלה ממשלה  
 ממשלה ממשלה ממשלה ממשלה ממשלה  
 ממשלה ממשלה ממשלה ממשלה ממשלה



(3) למה נשתמש ב-feature selection?  
כי לפעמים כמות המופעים שלנו יכולה להיות  
לחיצה למיליון וחמישים אלפי מופעים לכל יכולים  
להתחבר עם כמות כל כך גדולה של features.  
אנו צריכים בדרך כלל להוריד מנתן איתן  
של המוצר, כדי להוריד מנתן כיבוי, כדי  
להפחית מרעש בלוח ערכי מופעים לא קשורים  
ויש למצוא מניצת מוצר מופעים שלא יהיה  
מקנה לא נכונה.

למצוא בהרצאה שלוש סוגים של feature selection:  
Mutual Information, Chi-square, Frequency.

בשיטה Chi-square היא שיטה המבוססת  
על נחשים סטטיסטיים, ויכולה לזהות  
מופעים בלוח שימושים לפעולה סיווג.  
מחשבים על ידי הסתברות של הופעת מופעים  
מסוימים ביחס עם class מסוימים.

בשיטה Mutual Information היא שיטה המבוססת  
על מידע יותר מנצח ופרסומי מנצח. אבל  
יכולה להשתמש במופעים לא שימושים נצחים.  
מקבלים מוצר הכי גבוה את מופע המצא המסומן  
ומסומן שייך ל-class מסוימים.  
ושיטה הכי פשוטה היא Frequency. משהייה  
בה המופעים הם גבוהים. אין בה הסתמכות על  
זה עובד כי עוקבים עם גורמים הכי שימושים וברור  
המקרים זו שיטה הכי יעילה.

אנו נשתמש ב-feature selection ב-MB מופעים ונראות כי  
אחרת נסבוק מרעש בלוחים והרבה סבירות.



(4) למדנו בהרצאה על שיטות אחרות לסיווג  
 לקט והייתי מודע ורציתי להקל כ"ן שיטה  
 מודליות ומונחים (שאלתי). היענו למסקנה שהקיום  
 סיווג לקט אלו צריכים שיטה מודליות כפי  
 לקט מוצאנו יותר מודליות. בהנ"ל מודל אנו  
 מחפשים צורות שישו על יצי  $vector counts$   
 ויש אפשר צדק קוד-לז. ואם נחשב הסתברות  
 של הפער מופיע שישו במקום מודל"ה בכל  
 מסתק. אפשר שם צדק שיטה מודליות וריאנטי ולחזות  
 הלא מודל קיימת בכל מסתק אולי.



# סיכום

המערכת זאלא למצוא ביל' מ'הצ'ה

מספיק'ה המוכר'ה למ' כמ' למצוא א'ק

למ' מוצ' ס'ה'ל' de naive bayes

המצ'ה ו'ה ז'ק ספ'יו' פ'יו'ן ע'ו'ה

מ'צ'ה מוצ'ה ע'ו'ה . ז'כ'ס' ו'ה' ע'ו'ה'ק

מ'ה'ה א'ל'ט'ר'י'ט' de naive bayes ו'א'י'ה

ס'ו'י מ'י'ש'ו'ס ע'ו' ק"מ'ה' .

למצ'ה ע'ו'ה ע'ו'ה de feature selection

ו'מ'ה' כ'ל'א'ה' מ'ה'ן י'כ'ו'ל' ע'ו'ה' ע'ו'ה'ה' .

ל'נ' ו'א'י'ה' ס'ו' de naive bayes מ'ה'א'י'ה' .

המע'ה'ה זאלא ע'ו'ה'ה' ע'ו'ה'ה' ע'ו'ה'ה' .

מוצ'ה'ה' ו'ס'ו'ה' ס'ו'ה' ו'ה'ע'ה' ע'ו'ה'ה' ע'ו'ה'ה' .

מ'ו'ל'א'י'ו'מ'א'ל' ע'ו'ה' ס'ו'ה' ה'כ' מ'ה'א'י'ה' ו'י'ע'ה' .

ל'ס'ו'ה' ק'ס' . א'ל'ט'ר'י'ט' מ'ה'י'ס'ה' ל'כ'ה' ה'מ'י'ס'ה' .

ל'נ' כ'ל'ה' ה'מ'ס'ר'ה' ב'מ'י'ק'ו'ה' מ'ס'ו'ה' . ב'ל'ו' .

ק'ב'צ'ה' ל'א'י'ו'ן ו'ל'ק'ב'צ'ה' ו'ה'ע'ה'ה' ג'ו'ה'ה' .

ע'ו'ה' ה'מ'ו'ל'ה'ה' ו'כ'א'י'ו' ע'ו'ה'ה' ע'ו'ה'ה' .

ז'ו'ק' ה'כ' ע'ו'ה' . ו'ה' ע'ו'ה' ע'ו'ה' ע'ו'ה' .

כ'א'י'ו' ע'ו'ה' ע'ו'ה' . ו'ה' ע'ו'ה' . Multinomial .

א'ל'ט'ר'י'ט' ה'כ' מ'צ'ה' מ'כ'ה' ה'מ'ו'ל'ה'ה' .

ה'ה' .