מבוא ללמידה ממוכנת: שיעורי בית 1 –

מגישים:

איתי בר – 325839710

ליאון גורין – 214511214

שאלה 1:

נרצה לדעת אם לקנות או לא לקנות ולשם כך נרצה לגלות מי יותר גדול: $\Pr(buy\ TV)$ או $\Pr(don't\ buy\ TV)$. (החישוב נעשה בדומה להרצאה)

נרצה לחשב קודם את ה risk:

$$R(\alpha_i|x) = \sum_{j=1}^{c} \lambda(\alpha_i|w_j) \Pr(w_j|x) = \sum_{j=1}^{c} \lambda(\alpha_i|w_j) \cdot \frac{\Pr(x|w_j)\Pr(w_j)}{\Pr(x)}$$

נחשב את $\Pr(x)$ שמתאר במקרה שלנו את את איכות התמונה של הטלוויזיה:

$$\Pr(sharp\ picture)$$
$$= \Pr(sharp|good) \cdot \Pr(good) + \Pr(sharp|fair) \cdot \Pr(fair) + \Pr(sharp|bad) \cdot \Pr(bad)$$
$$= 0.9 \cdot 0.3 + 0.5 \cdot 0.5 + 0.2 \cdot 0.2 = 0.56$$

$$\Pr(diminished\ picture) = 1 - \Pr(sharp\ picture) = 0.44$$

נחשב את $\Pr(w_j|x) = \frac{\Pr(x|w_j)\Pr(w_j)}{\Pr(x)}$ לכל פרמוטציה של x, w_j:

$$\Pr(good\ TV|sharp\ picture) = \frac{\Pr(sharp|good) \cdot \Pr(good)}{\Pr(sharp)} = \frac{0.9 \cdot 0.3}{0.56} = 0.482$$

$$\Pr(bad\ TV|sharp\ picture) = \frac{\Pr(sharp|bad) \cdot \Pr(bad)}{\Pr(sharp)} = \frac{0.2 \cdot 0.2}{0.56} = 0.07$$

$$\Pr(fair\ TV|sharp\ picture) = 1 - \Pr(good\ TV|sharp\ picture) - \Pr(bad\ TV|sharp\ picture) = 0.448$$

$$\Pr(good\ TV|dim\ picture) = \frac{\Pr(dim|good) \cdot \Pr(good)}{\Pr(dim)} = \frac{0.1 \cdot 0.3}{0.44} = 0.07$$

$$\Pr(bad\ TV|dim\ picture) = \frac{\Pr(dim|bad) \cdot \Pr(bad)}{\Pr(dim)} = \frac{0.8 \cdot 0.2}{0.44} = 0.364$$

$$\Pr(fair\ TV|dim\ picture) = 1 - \Pr(good\ TV|dim\ picture) - \Pr(bad\ TV|dim\ picture) = 0.566$$

בעת נוכל לחשב $R(\alpha_i|x)$ לכל פרמוטציה ולהגיד מה ההחלטה האופטימלית לכל מצב.

$$R(buy\ TV|sharp\ picture) = \lambda(buy\ TV|good) \cdot \Pr(good|sharp\ picture)$$
$$+\lambda(buy\ TV|fair) \cdot \Pr(fair|sharp\ picutre)$$
$$+\lambda(buy\ TV|bad) \cdot \Pr(bad|sharp\ picutre)$$
$$= 0 \cdot 0.482 + 5 \cdot 0.448 + 20 \cdot 0.07 = 3.64$$

$$R(don't\ buy\ TV|sharp\ picture) = \lambda(don't\ buy\ TV|good) \cdot \Pr(good|sharp\ picture)$$
$$+\lambda(don't\ buy\ TV|fair) \cdot \Pr(fair|sharp\ picutre)$$
$$+\lambda(don't\ buy\ TV|bad) \cdot \Pr(bad|sharp\ picutre)$$
$$= 10 \cdot 0.482 + 5 \cdot 0.448 + 0 \cdot 0.07 = 7.26$$

זה אומר שיש סיכון גבוהה יותר לא לקנות, לכן אם התמונה של הטלוויזיה חדה נקנה!

$$R(buy\ TV|dim\ picture) = \lambda(buy\ TV|good) \cdot \Pr(good|dim\ picture)$$
$$+\lambda(buy\ TV|fair) \cdot \Pr(fair|dim\ picutre)$$
$$+\lambda(buy\ TV|bad) \cdot \Pr(bad|dim\ picutre)$$
$$= 0 \cdot 0.07 + 5 \cdot 0.566 + 20 \cdot 0.364 = 10.11$$

$$R(don't\ buy\ TV|dim\ picture) = \lambda(don't\ buy\ TV|good) \cdot \Pr(good|dim\ picture)$$
$$+\lambda(don't\ buy\ TV|fair) \cdot \Pr(fair|dim\ picutre)$$
$$+\lambda(don't\ buy\ TV|bad) \cdot \Pr(bad|dim\ picutre)$$
$$= 10 \cdot 0.07 + 5 \cdot 0.566 + 0 \cdot 0.364 = 3.53$$

זה אומר שיש סיכון גבוהה יותר לקנות, לכן אם התמונה של הטלוויזיה לא חדה לא נקנה!

ה risk הכולל יהיה:

$$total - risk = E_{\Pr(x)}[R(\alpha x|x)] = \Pr(sharp) \cdot \Pr(buy|sharp) + \Pr(dim) \cdot \Pr(don't\ buy|dim)$$

$$= 0.56 \cdot 3.64 + 0.44 \cdot 3.55 = 3.6004$$

שאלה 2:

א. נשתמש בחוק בייז בשביל לפתור את השאלה:

$$\Pr(have\ disease|tested\ positive) = \frac{\Pr(tested\ positive|have\ disease) \cdot \Pr(have\ disease)}{\Pr(tested\ positive)}$$

נתון:

$$\Pr(have\ disease) = 0.001$$

$$\Pr(tested\ negative|have\ disease) = 0$$

$$\Pr(tested\ positive|don't\ have\ disease) = 0.01$$

נחשב את החלקים החסרים:

$$\Pr(tested\ positive|have\ disease) = 1 - \Pr(tested\ negative|have\ disease) = 1$$

נסמן:

$$tested\ positive = pos, \qquad have\ disease = have, \qquad don't\ have\ disease = don't$$

$$\Pr(tested\ positive) = \Pr(pos|have) \cdot \Pr(have) + \Pr(pos|don't) \cdot \Pr(don't)$$

$$= 1 \cdot 0.001 + 0.01 \cdot (1 - 0.001) = 0.01099$$

נציב במשוואה המקורית:

$$\Pr(have\ disease|tested\ positive) = \frac{1 \cdot 0.001}{0.01099} = 0.09099181073$$

ב. אם הייתי הולך להתחסן, יש לי סיכוי גבוהה יותר שאין לי את המחלה כי הסיכוי שיש לי את המחלה הוא 0.001 לעומת סיכוי שאין לי את המחלה שהוא 0.999. תוצאת הבדיקה הוא משתנה בלתי תלוי להסתברות שיש לי או אין לי את המחלה.

ג. התשובה ל-(א) תהיה שונה כי MLE לא מתייחס ל posterior ולכן יחליט בא' בדיוק הפוך ממה שיצא לנו. התשובה ל-(ב) לא תהיה שונה כי אין התייחסות לשום posterior בחישוב שביצענו ונקבל את אותה התשובה

שאלה 4:

הסבר על הקוד –

פונקצית ()def learn_NB_text:

1. הקוד קולט את המידע על ה training data עם readTrainData
2. הקוד סופר כמה משפטים יש מכל קטגוריה לתוך category_occurrence_dict ומחשב $\Pr(category)$ ומכניס לתוך P לפי סדר ידוע (קבוע).
3. הקוד עובר על הקבצים וסופר מספר מופעמים של מילה מסוימת בקטגוריה ואת מספר המילים בקטגוריה.
4. לבסוף הוא עובר על כל המילים ומחשב את ההסתברות של מילה בהינתן קטגוריה (עם Laplacian smoothing)
$$\Pr\left(\frac{\#word-occurences-in-category+1}{\#word-in-categories+\#of\ categories}\right)$$
5. מכניסים את כל ההסתברויות של מילים כתלות בקטגוריה כרשימה לתוך Pw ומקבלים מטריצה (השורה הראשונה של Pw הגדרתי להיות רשימת מילים לפי הסדר)

פונקצית (def ClassifyNB_text(Pw, P:

1. הקוד מקבל את המידע על ה test data עם readTrainData
2. הקוד עובר על כל הקבצים לפי הסדר ועבור כל קובץ הוא מחשב את ההסתברות שהקובץ בקטגוריה מסוימת
   a. אנחנו בעצם מחשבים את:
   $$\log(Pr(category|text)) = \log\Pr(text|category) + \log(category)$$
   $$\log\Pr(text|category) = \log(word_1|category) + \cdots + \log(word_n|category)$$
   b. אנחנו מאתחלים את ההסתברות עבור הקטגוריה הנוכחית ומוסיפים log של כל הסתברויות של מילה כתלות בקטגוריה (אם המילה קיימת).
   c. אחרי שעוברים על כל הקטגוריות בוחרים בקטגוריה שקיבלה את ההסתברות הכי גבוהה ובודקים אם צדקנו. אם כן מוסיפים 1 להצלחות.
3. מחזירים את מספר ההצלחות חלקי מספר הקבצים.

הקוד בסוף מדפיס את ה success rate שחושב.

שאלה 3:

א)

נסמן הטלה של קובייה מספר 1, בתור

$$z_k^j = \begin{cases} 1 & if\ the\ k^{th}\ dice\ throw\ shows\ j \\ 0 & otherwise \end{cases}$$

לכן, $z_k^j \sim Ber(\theta)$, כאשר $\theta_j = P_1(j)$.

נחשב את פונקציית ה likelihood עבור θ:

$$L(\theta_j) = P\left(z_1^j, \ \dots, z_{40}^j | \theta_j\right) = \prod_{k=1}^{40} P\left(z_k^j | \theta_j\right) = \prod_{k=1}^{40} \theta_j^{z_k^j}\left(1 - \theta_j^{1-z_k^j}\right)$$

$$\ln L(\theta_j) = \sum_{k=1}^{40} z_k^j \cdot \ln(\theta_j) + \left(1 - z_k^j\right)\ln(1 - \theta_j)$$

$$\frac{\partial \ln L(\theta_j)}{\partial \theta_j} = \sum_{k=1}^{40} \frac{z_k^j}{\theta_j} - \frac{1 - z_k^j}{1 - \theta_j} = 0$$

$$\frac{1}{\theta}\sum_{k=1}^{40}\left(z_k^j\right) - \frac{40}{1 - \theta_j} + \frac{1}{1 - \theta_j}\sum_{k=1}^{40}\left(z_k^j\right) = 0$$

$$(1 - \theta_j)\sum_{k=1}^{40}\left(z_k^j\right) - 40\theta_j + \theta_j\sum_{k=1}^{40}\left(z_k^j\right) = 0_j$$

$$\sum_{k=1}^{40}\left(z_k^j\right) = 40\theta_j$$

$$\hat{\theta}_j = \frac{\sum_{k=1}^{40}\left(z_k^j\right)}{40} \approx \frac{\#\ throws\ showing\ j}{\#\ total\ throws}$$

למעשה, אם נבצע את אותו החישוב עבור כל אחד מצדדי הקובייה וגם עבור הקובייה השנייה, נקבל שה priors הם חלקיות ההטלות שהראו מספר מסוים מתוך מספר ההטלות הכולל. כלומר

$P_i(j) = \frac{\#\ throws\ showing\ j}{\#\ total\ throws}$.

על כן,

Dice 1:

$$P_1(1) = \frac{5}{40} = \frac{1}{8}$$
$$P_1(2) = \frac{3}{40} = \frac{3}{40}$$
$$P_1(3) = \frac{10}{40} = \frac{1}{4}$$
$$P_1(4) = \frac{1}{40} = \frac{1}{40}$$
$$P_1(5) = \frac{10}{40} = \frac{1}{4}$$
$$P_1(6) = \frac{11}{40} = \frac{11}{40}$$

Dice 2:

$$P_2(1) = \frac{10}{40} = \frac{1}{4}$$
$$P_2(2) = \frac{11}{40} = \frac{11}{40}$$
$$P_2(3) = \frac{4}{40} = \frac{1}{10}$$
$$P_2(4) = \frac{10}{40} = \frac{1}{4}$$
$$P_2(5) = \frac{3}{40} = \frac{3}{40}$$
$$P_2(6) = \frac{2}{40} = \frac{1}{20}$$

ב)

נרצה עכשיו לראות איזה קובייה בעלת הסתברות גבוהה יותר בהינתן ההטלות החדשות. כלומר נרצה להסתכל על

$$P(D_1|x_1, \dots, x_{40})$$
$$P(D_2|x_1, \dots, x_{40})$$

נשתמש בכלל bayes:

$$P(D_1|x_1, \dots, x_{40}) = \frac{P(x_1, \dots, x_{40}|D_1)P(D_1)}{P(x_1, \dots, x_{40})}$$

מאחר ולא נתון שלאחת מהקוביות יש הסתברות גבוהה יותר להיעלם נוכל להתעלם מ $P(D_1)$, ובנוסף גם מ

$P(x_1, \ldots, x_{40})$, מפני שאנחנו מתעניינים רק בהשוואה בין שתי הקוביות.

$$P(x_1, \ldots, x_{40}|D_1) = P(x_1|D_1) \cdot \ldots \cdot P(x_{40}|D_1)$$
$$= \left(\frac{1}{8}\right)^8 \cdot \left(\frac{3}{40}\right)^{12} \cdot \left(\frac{1}{4}\right)^6 \cdot \left(\frac{1}{40}\right)^9 \cdot \left(\frac{1}{4}\right)^4 \cdot \frac{11}{40} = 1.88 \cdot 10^{-42}$$

נחשב גם עבור ההסתברות לקובייה השנייה באותו אופן:

$$P(x_1, \ldots, x_{40}|D_2) = P(x_1|D_2) \cdot \ldots \cdot P(x_{40}|D_2)$$
$$= \left(\frac{1}{4}\right)^8 \cdot \left(\frac{11}{40}\right)^{12} \cdot \left(\frac{1}{10}\right)^6 \cdot \left(\frac{1}{4}\right)^9 \cdot \left(\frac{3}{40}\right)^4 \cdot \frac{1}{20} = 1.72 \cdot 10^{-29}$$

קיבלנו ש

$$P(x_1, \ldots, x_{40}|D_1) < P(x_1, \ldots, x_{40}|D_2)$$

ולכן קובייה 2 היא זו שסבירה יותר שתביא לתוצאות שקיבלנו. כלומר הקובייה שאיבדנו היא קובייה 1.

שאלה 5

מוגדרת עבורנו פונקציית loss 0-1:

$$\lambda(a_i|c_j) = \begin{cases} 1 & i \neq j \\ 0 & i = j \end{cases}$$

הסיכון המותנה מוגדר:

$$R(a_i|x) = \sum_j P(c_j|x) \cdot \lambda(a_i|c_j) = 1 - P(c_i|x)$$

על פי bayesian decision rule, נרצה לבחור את ההחלטה שעבורה הסיכון המותנה הוא הקטן ביותר:

$$i = \min R(a_i|x) = \min 1 - P(c_i|x) = \max P(c_i|x)$$

למעשה קיבלנו מסווג MAP, מפני שאנו בוחרים את המחלקה שעבורה ההסתברות היא הכי גבוהה בהינתן המאפיינים.