מבוא ללמידת מכונה להנדסת תוכנה

עבודת הגשה 3 – Bayesian Learning and Ensemble

תאריך ההגשה (דרך המודל): 23:59 13/08/2024

שאלה 1

מרצה הקורס "מבוא ללמידת מכונה" שמלמד אותו מזה מספר שנים מעוניין להבין את הגורמים להצלחת הסטודנטים בקורס.

לשם כך, מרצה הקורס אסף מידע לגבי מספר גורמים הקשורים להתנהגות הסטודנטים בקורס לצד ביצועיהם הסופיים.

הדאטאסט כולל את העמודות הבאות:

1. Record – מספר רשומה.
2. StudyHours- משך הזמן שהסטודנט השקיע בלמידה (High, Medium, Low).
3. Attendance – נוכחות בשיעורים (High, Medium, Low).
4. Preparation – איכות ההכנה למבחן (High, Medium, Low).
5. PassedExam – האם הסטודנט עבר את הבחינה (Yes, No).

עזרו למרצה הקורס לנבא את הצלחת הסטודנטים לעבור את הבחינה הסופית בקורס באמצעות אלגוריתם Naïve Bayes ובנתונים שבטבלה הבאה כדי לחזות האם שלושת הסטודנטים האחרונים ברשימה יעברו או שלא יעברו את הבחינה.

לשם כך, בנו את מודל Naïve Bayes על סמך סט האימון (Training Set) הכולל רשומות 1-12 כאשר סט המבחן (Test Set) הן הרשומות מ-13 עד 15.

א. בצעו חישובים מלאים תוך שימוש ב-Laplacian estimate.

ב. הפעילו את המסווג שהשתמש ב-Laplacian estimate על ה-Test Set. מה הדיוק שקיבלתם? מי הסטודנטים שעברו ולא עברו את הבחינה?

ג. בצעו חישובים מלאים תוך שימוש ב-m-estimate. $m=3$

ד. הפעילו את המסווג שהשתמש ב-m-estimate על ה-Test Set. מה הדיוק שקיבלתם? מי הסטודנטים שעברו ולא עברו את הבחינה?

Record	StudyHours	Attendance	Preparation	PassedExam
1	High	High	Good	Yes
2	High	High	Average	Yes
3	Medium	High	Good	Yes
4	Low	High	Good	Yes
5	Low	Low	Poor	No
6	Medium	Low	Average	No
7	Medium	Low	Poor	No
8	Low	High	Average	No
9	High	Low	Good	Yes
10	High	Low	Poor	No
11	Medium	High	Average	Yes
12	Medium	High	Poor	No
13	Low	High	Poor	No
14	High	High	Poor	No
15	High	Low	Average	Yes

א) $P(\text{passedExam} = \text{yes}) = \frac{7+1}{12+2} = \frac{8}{14} = 0.571$

$P(\text{passedExam} = \text{No}) = \frac{5+1}{12+2} = \frac{6}{14} = 0.429$

Study Hours :

$P(\text{Study Hours} = \text{Hight} \mid \text{Passed Exam} = \text{yes}) = \frac{3+1}{7+3} = 0.4$

$P(\text{Study Hours} = \text{Meduim} \mid \text{Passed Exam} = \text{yes}) = \frac{2+1}{7+3} = 0.3$

$P(\text{Study Hours} = \text{Low} \mid \text{Passed Exam} = \text{yes}) = \frac{1+1}{7+3} = 0.2$

$P(\text{Study Hours} = \text{Hight} \mid \text{Passed Exam} = \text{No}) = \frac{1+1}{5+3} = \frac{2}{8} = 0.25$

$P(\text{Study Hours} = \text{Meduim} \mid \text{Passed Exam} = \text{No}) \quad \frac{2+1}{5+3} = \frac{3}{8} = 0.375$

$P(\text{Study Hours} = \text{Low} \mid \text{Passed Exam} = \text{No}) = \frac{2+1}{5+} = \frac{3}{8} = 0.375$

Attendance:

$p(\text{Attendance} = \text{Hight} \mid \text{passExam} = \text{yes}) = \frac{5+1}{7+2} = 0.67$

$p(\text{Attendance} = \text{Low} \mid \text{passedExam} = \text{yes}) = \frac{2+1}{7+2} = 0.33$

$p(\text{Attendance} = \text{Hight} \mid \text{passExam} = \text{No}) = \frac{3+1}{5+2} = 0.57$

$p(\text{Attendance} = \text{Low} \mid \text{passedExam} = \text{No}) = \frac{2+1}{5+2} = 0.42$

preperation:

$p(\text{preperation} = \text{Good} \mid \text{passedExam} = \text{yes}) = \frac{3+1}{7+3} = 0.4$

$p(\text{preperation} = \text{Average} \mid \text{passedExam} = \text{yes}) = \frac{2+1}{7+3} = 0.3$

$p(\text{Preperation} = \text{poor} \mid \text{passedExam} = \text{yes}) = \frac{2+1}{7+3} = 0.3$

$p(\text{preperation} = \text{Good} \mid \text{passedExam} = \text{No}) = \frac{1+1}{5+3} = \frac{2}{8} = 0.$

$p(\text{preperation} = \text{Average} \mid \text{passedExam} = \text{No}) = \frac{2+1}{5+3} = \frac{3}{8} = 0.3$

$p(\text{Preperation} = \text{poor} \mid \text{passedExam} = \text{No}) = \frac{3+1}{5+3} = \frac{3}{8} = 0.3$

Record 13:

$P(\text{passedExam} = \text{yes} \mid \text{Record 13})$:

$P(\text{passedExam} = \text{yes}) \cdot P(\text{studyHours} = \text{low} \mid \text{passedExam} = \text{yes}) \cdot P(\text{Attendance} \mid$

$= 0.571 \cdot 0.2 \cdot 0.67 \cdot 0.3 = 0.0229542$

$P(\text{passedExam} = \text{No})$

גדול יותר

$0.43 \cdot 0.375 \cdot 0.57 \cdot 0.375 = 0.0344$

יותר טוב

Record 14:

$P(\text{passedExam} = \text{yes} \mid \text{Record 14}) =$

$= 0.571 \cdot 0.4 \cdot 0.67 \cdot 0.3 = 0.0459$

$P(\text{passedExam} = \text{No} \mid \text{Record 14}) =$

$0.43 \cdot 0.25 \cdot 0.57 \cdot 0.375 = 0.022$

יותר

Record 15:

P(PassedExam = yes / Record 15):

$0.571 \cdot 0.4 \cdot 0.33 \cdot 0.3 = 0.0226$

P(PassedExam = No / Record 15):

$0.43 \cdot 0.25 \cdot 0.43 \cdot 0.375 = 0.01733$

Record 13 : peredict No Actual NO

Record 14 : perdict yes Actual No

Record 15 : perdict yes Actual yes

$$\text{Accuracy} = \frac{\text{Num of correct prediction}}{\text{total Num of predection}} =$$

$$P(\text{studyHourse} = \text{Hight} \mid \text{passedExam} = \text{yes}) = \frac{\frac{3}{12} \cdot 3 + 3}{7 + 3} =$$

ג)

חישוב ההסתברויות המותנות בעזרת תיקון לפלס עם $m=3$.

$$P(PassedExam=yes)=\frac{7+3}{12+6}=\frac{10}{18}=0.56$$

$$P(PassedExam=No)=\frac{5+3}{12+6}=\frac{8}{18}=0.44$$

StudyHours:

$$P(StudyHours=Hight \mid PassedExam=yes)=\frac{0.4\cdot 3+4}{10}=0.4$$

$$P(StudyHours=Meduim \mid PassedExam=yes)=\frac{0.3\cdot 3+3}{10}=0.3$$

$$P(StudyHours=Low \mid PassedExam=yes)=\frac{0.3\cdot 1+3}{10}=0.2$$

$$P(StudyHours=Hight \mid PassedExam=No)=\frac{0.3\cdot 1+3}{8}=0.25$$

$$P(StudyHours=Meduim \mid PassedExam=No)=\frac{0.3\cdot 2+3}{8}=0.375$$

$$P(StudyHours=Low \mid PassedExam=No)=\frac{0.3\cdot 2+3}{8}=0.375$$

Attendance:

$P(\text{Attendance} = \text{Hight} \mid \text{passExam} = \text{yes}) = \frac{5+1.5}{7+2} = 0.67$

$P(\text{Attendance} = \text{Low} \mid \text{passedExam} = \text{yes}) = \frac{2+1}{7+2} = 0.33$

$P(\text{Attendance} = \text{Hight} \mid \text{passExam} = \text{No}) = \frac{3+1}{5+2} = 0.57$

$P(\text{Attendance} = \text{Low} \mid \text{passedExam} = \text{No}) = \frac{2+1}{5+2} = 0.43$

preperation:

$P(\text{preperation} = \text{Good} \mid \text{passedExam} = \text{yes}) = \frac{3+1}{7+3} = 0.4$

$P(\text{preperation} = \text{Average} \mid \text{passedExam} = \text{yes}) = \frac{2+1}{7+3} = 0.4$

$P(\text{Preperation} = \text{poor} \mid \text{passedExam} = \text{yes}) = \frac{2+1}{7+3} = 0.2$

$P(\text{preperation} = \text{Good} \mid \text{passedExam} = \text{No}) = \frac{1+1}{5+3} = \frac{2}{8} = 0.25$

$P(\text{preperation} = \text{Average} \mid \text{passedExam} = \text{No}) = \frac{3+1}{5+3} = \frac{4}{8} = 0.5$

$P(\text{Preperation} = \text{poor} \mid \text{passedExam} = \text{No}) = \frac{1+1}{5+3} = \frac{2}{8} = 0.25$

טעות ב 13 !

13) yes $0.2 \cdot 0.6 \cdot 0.2 \cdot 0.56 = 0.0148$

לא $0.25 \cdot 0.56 \cdot 0.375 \cdot 0.43 = 0.0229$

לא עבר X

לא
לא
לא

14)

טעות ב 14 :-

yes $0.2 \cdot 0.65 \cdot 0.4 \cdot 0.56 = 0.0295$

לא $0.25 \cdot 0.56 \cdot 0.25 \cdot 0.43 = 0.0153$

עבר ✓

טעות ב 15 :

yes $0.4 \cdot 0.3 \cdot 0.4 \cdot 0.56 = 0.03$

לא $0.5 \cdot 0.43 \cdot 0.25 \cdot 0.43 = 0.02$

עבר ✓

ציון : $\frac{2}{3} = 67\%$

סיום :

2) $p(\text{passedExam} = \text{yes}) = \frac{7+3}{12+6} = 0.56$

$p(\text{passedExam} = \text{No}) = \frac{5+3}{12+6} = 0.44$

Study Hours:

$p(\text{studyHours} = \text{Hight} \mid \text{Pased Exam} = \text{yes}) = \frac{3+1}{7+3} = 0.4$

$p(\text{studyHours} = \text{Meduim} \mid \text{passedExam} = \text{yes}) = \frac{2+1}{7+3} = 0.3$

$p(\text{studyHours} = \text{Low} \mid \text{PassedExam} = \text{yes}) = \frac{2+1}{7+3} = 0.3$

$p(\text{studyHours} = \text{Hight} \mid \text{Pased Exam} = \text{No}) = \frac{1+1}{5+3} = \frac{2}{8} = 0.25$

$p(\text{studyHours} = \text{Meduim} \mid \text{passedExam} = \text{No}) \ \frac{2+1}{5+3} = \frac{3}{8} = 0.375$

$p(\text{studyHours} = \text{Low} \mid \text{PassedExam} = \text{No}) = \frac{2+1}{5+} = \frac{3}{8} = 0.375$

Attendance:

$p(\text{Attendance} = \text{Hight} \mid \text{PassExam} = \text{yes}) = \frac{5+1}{7+2} = 0.67$

$p(\text{Attendance} = \text{Low} \mid \text{passedExam} = \text{yes}) = \frac{2+1}{7+2} = 0.33$

$P(\text{Attendance} = \text{Hight} \mid \text{passExam} = \text{No}) = \frac{3+1}{5+2} = 0.57$

$P(\text{Attendance} = \text{Low} \mid \text{passedExam} = \text{No}) = \frac{2+1}{5+2} = 0.43$

preperation:

$P(\text{preperation} = \text{Good} \mid \text{passedExam} = \text{yes}) = \frac{3+1}{7+3} = 0.4$

$P(\text{preperation} = \text{Average} \mid \text{passedExam} = \text{yes}) = \frac{2+1}{7+3} = 0.3$

$P(\text{Preperation} = \text{poor} \mid \text{passedExam} = \text{yes}) = \frac{2+1}{7+3} = 0.3$

$P(\text{preperation} = \text{Good} \mid \text{passedExam} = \text{No}) = \frac{1+1}{5+3} = \frac{2}{8} = 0.25$

$P(\text{preperation} = \text{Average} \mid \text{passedExam} = \text{No}) = \frac{2+1}{5+3} = \frac{3}{8} = 0.375$

$P(\text{Preperation} = \text{poor} \mid \text{passedExam} = \text{No}) = \frac{3+1}{5+3} = \frac{3}{8} = 0.375$

- בפעמה שסעיף ב בונים שימוש ב m-estimate ונשתמש בסיומת חישובים בנ"ל בלוק: $\frac{2}{3}$