

WEEK 7 LAB

Naives Bayes Tutorial

Task 1: Accident Prediction For the table above, find the probability of the scenario being in an accident (YES or NO).

1. Objective: Calculate $P(C1 | X = (\text{Rain, Good, Normal, No}))$. Determine how the Bayes classifier would classify the data instance $X = (\text{Rain, Good, Normal, No})$.

Total instance = 10

YES = 5

NO = 5

$$P(\text{accident} = \text{yes}) = 5/10 = 0.5$$

$$P(\text{accident} = \text{no}) = 5/10 = 0.5$$

Calculating probabilities when accident = yes where $X = (\text{rain, good, normal, no})$

$$\text{For rain; } P(\text{rain/accident} = \text{yes}) = 1/5 = 0.2$$

$$\text{For good; } p(\text{good/accident=yes}) = 1/5 = 0.2$$

$$\text{For normal; } p(\text{normal/accident=yes}) = 1/5 = 0.2$$

$$\text{For No; } p(\text{no/accident=yes}) = 2/5 = 0.4$$

Calculating probabilities when accident = No where $X = (\text{rain, good, normal, no})$

$$\text{For rain; } P(\text{rain/accident} = \text{no}) = 2/5 = 0.4$$

$$\text{For good; } p(\text{good/accident=yes}) = 3/5 = 0.6$$

$$\text{For normal; } p(\text{normal/accident=yes}) = 2/5 = 0.4$$

$$\text{For No; } p(\text{no/accident=yes}) = 4/5 = 0.8$$

Calculating probability of X given that accident = yes,

$$P(X / \text{accident} = \text{yes}) = 0.2 * 0.2 * 0.2 * 0.4 = 0.0032$$

Calculating probability of X given that accident = No,

$$P(X / \text{accident} = \text{no}) = 0.4 * 0.6 * 0.4 * 0.8 = 0.0768$$

Now,

$$P(\text{accident} = \text{yes}/X) = 0.0032 * 0.5 = 0.0016$$

$$P(\text{accident} = \text{no}/X) = 0.0768 * 0.5 = 0.0384$$

Task 2: Weather-Based Game Prediction In this dataset, there are five categorical attributes: outlook, temperature, humidity, windy, and play. We are interested in building a system to classify whether to play based on weather conditions.

1. Question 1: Calculate $P(C_1 | X = (\text{sunny, hot, high, false}))$. How would the Bayes classifier classify the data instance $X = (\text{sunny, hot, high, false})$?

Total instance = 14,

Yes = 9

No = 5

$$P(\text{yes}) = 9/14 = 0.64$$

$$P(\text{no}) = 5/14 = 0.36$$

Calculating probabilities for X ($\text{sunny, hot, high, false}$) given that $\text{play} = \text{yes}$

$$P(\text{sunny}/\text{play} = \text{yes}) = 2/9 = 0.2$$

$$P(\text{hot}/ \text{play} = \text{yes}) = 2/9 = 0.2$$

$$P(\text{high} / \text{play} = \text{yes}) = 3/9 = 0.3$$

$$P(\text{false} / \text{play} = \text{yes}) = 6/9 = 0.67$$

Calculating probabilities for X ($\text{sunny, hot, high, false}$) given that $\text{play} = \text{No}$

$$P(\text{sunny}/\text{play} = \text{no}) = 3/5 = 0.6$$

$$P(\text{hot} / \text{play} = \text{no}) = 2/5 = 0.4$$

$$P(\text{high} / \text{play} = \text{no}) = 4/5 = 0.8$$

$$P(\text{false} / \text{play} = \text{no}) = 2/5 = 0.4$$

Calculating probability of X given that $\text{play} = \text{yes}$ and $\text{play} = \text{no}$

$$P(X/\text{play} = \text{yes}) = 0.2 * 0.2 * 0.3 * 0.67 = 0.00804$$

$$P(X/\text{play} = \text{no}) = 0.6 * 0.4 * 0.8 * 0.4 = 0.0768$$

Calculating probabilities of $\text{play} = \text{yes}$ and $\text{play} = \text{no}$ given that X

$$P(\text{play} = \text{yes}/X) = 0.00804 * 0.64 = 0.0051$$

$$P(\text{play} = \text{no}/X) = 0.0768 * 0.36 = 0.028$$

2. Question 2: Does this agree with the classification in Table 1 for $X = (\text{sunny, hot, high, false})$?
=> NO

3. Question 3: Consider a new data instance $X' = (\text{overcast}, \text{cool}, \text{high}, \text{true})$. How would the Bayes classifier classify X' ?

Total instance = 14,

Yes = 9

No = 5

$$P(\text{yes}) = 9/14 = 0.64$$

$$P(\text{no}) = 5/14 = 0.36$$

Calculating probabilities for X (overcast, cool, high, True) given that play = yes

$$P(\text{overcast}/\text{play} = \text{yes}) = 4/9 = 0.44$$

$$P(\text{cool}/ \text{play}=\text{yes}) = 3/9 = 0.33$$

$$P(\text{high} / \text{play}=\text{yes}) = 3/9 = 0.33$$

$$P(\text{true}/ \text{play}=\text{yes}) = 3/9 = 0.33$$

Calculating probabilities for X (overcast, cool, high, True) given that play = No

$$P(\text{overcast}/\text{play} = \text{no}) = 0/5 = 0$$

$$P(\text{cool}/ \text{play}=\text{no}) = 1/5 = 0.2$$

$$P(\text{high} / \text{play}=\text{no}) = 4/5 = 0.8$$

$$P(\text{true}/ \text{play}=\text{no}) = 3/5 = 0.6$$

Calculating probability of X given that play = yes and play = no

$$P(X/\text{play}=\text{yes}) = 0.44 * 0.33 * 0.33 * 0.33 = 0.0158$$

$$P(X/\text{play}=\text{no}) = 0 * 0.2 * 0.8 * 0.6 = 0$$

Calculating probabilities of play=yes and play=no given that X

$$P(\text{play}=\text{yes}/X) = 0.0158 * 0.64 = 0.0051$$

$$P(\text{play}=\text{no}/X) = 0 * 0.36 = 0$$

Task 3:

Loan Approval Prediction Use Naive Bayes to classify whether a new applicant's loan will be approved based on employment status, credit history, and income level.

1. A new applicant is Employed with a Good credit history and a Medium income level.

Objective: Calculate $P(\text{LoanApproved} = \text{Yes} | \text{EmploymentStatus} = \text{Employed}, \text{CreditHistory} = \text{Good}, \text{IncomeLevel} = \text{Medium})$ using Naive Bayes.

Total instance = 5

Yes = 3,

No= 2

$$P(\text{LoanApproved}=\text{Yes}) = 3/5 = 0.6$$

$$P(\text{LoanApproved}=\text{No}) = 2/5 = 0.4$$

Calculating probabilities for X(employed, good, medium) given that loanapproved = yes

$$P(\text{employed} / \text{yes}) = 2/3 = 0.67$$

$$P(\text{good} / \text{yes}) = 3/3 = 1$$

$$P(\text{medium} / \text{yes}) = 1/3 = 0.33$$

Calculating probabilities for X(employed, good, medium) given that loanapproved = no

$$P(\text{employed} / \text{no}) = 1/2 = 0.5$$

$$P(\text{good} / \text{no}) = 0/2 = 0$$

$$P(\text{medium} / \text{no}) = 1/2 = 0.5$$

$$P(X / \text{yes}) = 0.67 * 1 * 0.33 = 0.22$$

$$P(X / \text{no}) = 0.5 * 0.5 * 0 = 0$$

2. Another applicant is Unemployed with a Bad credit history and a Low income level.

Objective: Calculate $P(\text{LoanApproved} = \text{No} \mid \text{EmploymentStatus} = \text{Unemployed}, \text{CreditHistory} = \text{Bad}, \text{IncomeLevel} = \text{Low})$

For loan approved = no

$$P(\text{unemployed/no}) = 1/2 = 0.5$$

$$P(\text{bad} / \text{no}) = 2/2 = 1$$

$$P(\text{low} / \text{no}) = 1/2 = 0.5$$

For loan approved = yes

$$P(\text{unemployed/yes}) = 1/2 = 0.5$$

$$P(\text{bad} / \text{yes}) = 0/2 = 0$$

$$P(\text{low} / \text{yes}) = 1/2 = 0.5$$

$$P(X / \text{no}) = 0.5 * 0.5 * 1 = 0.25$$

$$P(X / \text{YES}) = 0.5 * 0 * 0.5 = 0$$

Here, probability of getting loanapproved= yes for (unemployment, bad, low) is less than that of loanapproved = no so the loan is not approved.

3. (Advanced) The bank introduces a scoring system for applicants, assigning scores to each feature (e.g., "Employed" has a score of 3, "Unemployed" has 1). Discuss how this might impact probability calculations.

Objective: Consider the impact of a scoring system on Naive Bayes probability calculations.

Task 4

Total instance = 5

Positive = 3

Negative = 2

$$P(\text{positive}) = 3/5 = 0.6$$

$$P(\text{negative}) = 2/5 = 0.4$$

FOR question 1:

For X (fever=yes, cough=no, fatigue=yes, travelhistory=no) given that diseasesdiagnosis = positive;

$$P(\text{fever} / \text{positive}) = 3/3 = 1$$

$$P(\text{cough} / \text{positive}) = 2/3 = 0.67$$

$$P(\text{fatigue} / \text{positive}) = 3/3 = 1$$

$$P(\text{travelhistory} / \text{positive}) = 2/3 = 0.67$$

Calculating probability of X when diseasesdiagnosis is positive

$$P(x/\text{positive}) = 1 * 0.67 * 1 * 0.67 = 0.45$$

Calculating probability of diseasesdiagnosis is positive given that of X

$$P(\text{positive} / X) = 0.45 * 0.6 = 0.27$$

For question 2:

For X (fever=no, cough=yes, fatigue=no, travelhistory=no) given that diseasesdiagnosis = negative;

$$P(\text{fever} / \text{negative}) = 2/2 = 1$$

$$P(\text{cough} / \text{negative}) = 2/2 = 1$$

$$P(\text{fatigue} / \text{negative}) = 2/2 = 1$$

$$P(\text{travelhistory} / \text{negative}) = 1/2 = 0.5$$

Calculating probability of X when diseasesdiagnosis is negative

$$P(x/\text{positive}) = 1 * 1 * 1 * 0.5 = 0.5$$

Calculating probability of diseases diagnosis is negative given that of X

$$P(\text{positive} / X) = 0.5 * 0.4 = 0.2$$