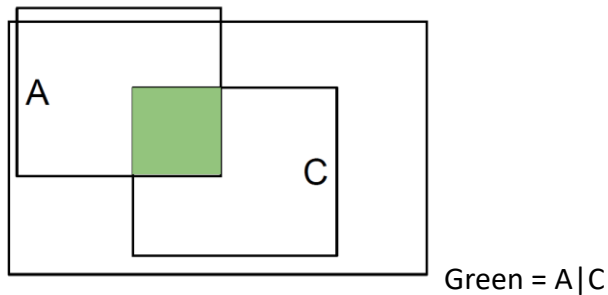# Lecture 13 B

Naive Bayes

Foundation:

Conditional Probability: $P(A|C) = \frac{P(A \cap C)}{P(C)}$

Green = A|C

Bayes Theorem:

$$P(A|C) = \frac{P(C|A)P(A)}{P(C)}$$

Example:

Given:

- Meningitis causes a stiff neck 50% of the time
- Prior probability of any patient having meningitis is $\frac{1}{50,000}$
- Prior probability of any patient having a stiff neck is $\frac{1}{20}$

If a patient has a stiff neck, what is the probability that they have meningitis?

$$P(M|S) = \frac{P(S|M)P(M)}{P(S)}$$

$$P(M|S) = \frac{0.5 * \frac{1}{50000}}{\frac{1}{20}}$$

$$P(M|S) = \frac{\frac{1}{100000}}{\frac{1}{20}} = 0.0002 \ OR \ 0.02\%$$

Bayesian Classifiers

A Bayesian Classifier applies Bayes' theorem for classification tasks.

It uses the posterior probability to decide the class label for new input data.

The classifier predicts the class with the highest posterior probability given the observed features.

Bayes theorem

$$P(C_k|X) = \frac{P(X|C_k)P(C_k)}{P(X)}$$

Where:

- $P(C_k|X)$ = posterior probability of class C_k given features X
- $P(X|C_k)$ = likelihood of features given class.
- $P(C_k)$ = prior probability of class C_k.
- $P(X)$ = probability of features (normalizing constant).

Working:

1. Estimate class priors: Probability of each class in the training data.
2. Estimate likelihood: Probability of feature values given class.
3. Apply Bayes' Theorem to compute posterior probabilities.
4. Classify new data as the class with the highest posterior.

Example: Predict whether a person will buy a computer (Yes or No) based on two features:

Age	Income	Buy Computer
Young	High	No
Young	High	No
Middle-aged	High	Yes
Senior	Medium	Yes
Senior	Low	Yes
Senior	Low	No
Middle-aged	Low	Yes
Young	Medium	No
Young	Low	Yes
Senior	Medium	Yes
Young	Medium	Yes
Middle-aged	Medium	Yes
Middle-aged	High	Yes
Senior	Medium	No

- Total samples = 14
- Number of Yes = 9
- Number of No = 5

$$P(Buy = yes) = \frac{9}{14} = 0.643 \; ; P(Buy = no) = \frac{5}{14} = 0.357$$

Calculate Likelihoods: $P(Age|Buy)\,, P(Income|Buy)$

For Buy = yes

Age	Count	Probability
Young	2	2/9 ≈ 0.222
Middle-aged	4	4/9 ≈ 0.444
Senior	3	3/9 ≈ 0.333

Income	Count	Probability
High	3	3/9 ≈ 0.333
Medium	4	4/9 ≈ 0.444
Low	2	2/9 ≈ 0.222

For Buy = No

Age	Count	Probability
Young	3	3/5 = 0.6
Middle-aged	0	0/5 = 0
Senior	2	2/5 = 0.4

Income	Count	Probability
High	2	2/5 = 0.4
Medium	2	2/5 = 0.4
Low	1	1/5 = 0.2

Predict for new sample: Age = Young, Income = Medium

Calculate posterior probabilities (using Naive Bayes independence assumption):

$$P(Yes|Young, Medium) = P(yes) * P(Young|yes) * P(Medium|yes)$$

$$= 0.643 * 0.222 * 0.444 = \mathbf{0.0633}$$

$$P(No|Young, Medium) = P(No) * P(Young|No) * P(Medium|No)$$

$$= 0.357 * 0.6 * 0.4 = \mathbf{0.0857}$$

Since $P(No|Young, Medium) > P(Yes|Young, Medium)$, Predict **_No_**

Refund	Marital Status	Income	Class
Yes	Single	125k	No
No	Married	100k	No
No	Single	70k	No
Yes	Married	120k	No
No	Divorced	90k	Yes
No	Married	60k	No
Yes	Divorced	220k	No
No	Single	85k	Yes
No	Married	75k	No
No	Single	90k	Yes

With Gaussian

$$Mean\ (No) = \frac{125+100+70+120+60+220+75}{7} = \frac{770}{7} = \mathbf{110}$$

$$Var\ (No) = \frac{(125-110)^2+(100-110)^2+(70-110)^2+(120-110)^2+(60-110)^2+(220-110)^2+(75-110)^2}{7}$$

$$= \frac{15^2+(-10)^2+(-40)^2+10^2+(-50)^2+110^2+(-35)^2}{7} = \frac{225+100+1600+100+2500+12100+1225}{7} = \frac{17850}{7} = \mathbf{2975}$$

$$\sigma^2 = Var(No)$$

Assumption: The continuous feature Income given class No follows a Normal (Gaussian) distribution.

Use Gaussian Probability Density Function (PDF):

$$P(x \mid C) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

For $x = 120k\ ; C = No$

$$P(Income = 120k|No) = \frac{1}{\sqrt{2\pi * 2975}} e^{-\left(\frac{(120-110)^2}{2*2975)}\right)}$$

$$= \frac{1}{\sqrt{2\pi * 2975}} e^{-0.0168} = \frac{1}{136.55} * 0.9833 = \mathbf{0.0072}$$

The value $P(Income = 120k|No) = 0.0072$ is the likelihood of seeing income 120k among samples in class No under the normal distribution assumption.

This likelihood is used in Naive Bayes to calculate posterior class probabilities for new data points with continuous features.

Without Gaussian

Discretize Income into bins

- Bin 1: 60k – <90k
- Bin 2: 90k – <120k
- Bin 3: 120k – <150k
- Bin 4: 150k – 180k+

Assign Income values for class No to bins

Income	Bin
125k	Bin 3
100k	Bin 2
70k	Bin 1
120k	Bin 3
60k	Bin 1
220k	Bin 4 (assuming upper bin includes all higher incomes)
75k	Bin 1

Therefore, samples per bin:

- Bin 1: 3
- Bin 2: 1
- Bin 3: 2
- Bin 4: 1

Total: 7

Estimate conditional probability from frequencies

$$P(Income = 120k|No) = \frac{Count\ in\ Bin\ 3}{Total\ Samples\ in\ Class\ No.} = \frac{2}{7} = \mathbf{0.286}$$

So, the probability of Income being in the 120k range (Bin 3) given class No is approximately 0.286.

For both $P(X|No)$ and $P(X|Yes)$

Naive Bayes assumption:

1. $P(X|No) = P(Refund = No|No) * P(Married|No) * P(Income = 120k|No)$
2. $P(X|Yes) = P(Refund = No|Yes) * P(Married|Yes) * P(Income = 120k|Yes)$

<u>For First:</u> $P(X|No)$

Count of Refund = No in class No (from data):

Refund = No and class No = 4 (rows 2,3,6,9)

Total class No samples = 7

$$P(Refund = No|No) = \frac{4}{7} = 0.571$$

Count Married in class No:

Married and class No = 4 (rows 2,4,6,9)

Total class No samples = 7

$$P(Married|No) = \frac{4}{7} = 0.571$$

For Income = 120k given class No

I. Using Gaussian (From previous info):

$$P(Income = 120k|No) = 0.0072$$

II. Using Binning (From Previous Info):
$$P(Income = 120k|No) = 0.286$$

Therefore, $P(X|No)$

I. For Gaussian: $P(X|No) = 0.571 * 0.571 * 0.0072 = 0.00234$
II. For Binning: $P(X|No) = 0.571 * 0.571 * 0.286 = 0.09324$

<u>For Second:</u> $P(X|Yes)$

Count of Refund = No in class Yes (from data):

Refund = No and class Yes = 3

Total class No samples = 7

$$P(Refund = No|Yes) = \frac{3}{7} = 0.429$$

Count Married in class Yes:

Married and class Yes = 1

Total class No samples = 7

$$P(Married|Yes) = \frac{1}{7} = 0.143$$

For Income = 120k given class Yes

    I.   Using Gaussian (Calculate using above method):

$$P(Income = 120k|Yes) = 0.0019$$

    II.   Using Binning (Calculate using above method):
$$P(Income = 120k|Yes) = 0.429$$

Therefore, $P(X|Yes)$

    I.   For Gaussian: $P(X|Yes) = 0.429 * 0.143 * 0.0019 = 0.0001165$
    II.   For Binning: $P(X|Yes) = 0.429 * 0.143 * 0.429 = 0.0263$

Final Tables for $P(X|No)$ and $P(X|Yes)$

Values	$P(X|No)$	$P(X|Yes)$	Prediction
Gaussian	0.00234	0.0001165	_No_
Binning	0.09324	0.0263	_No_

Interpretation:

1. Gaussian gives a probability density, which is usually quite low for specific points but useful when combining continuous attributes.
2. Binning gives a probability mass for intervals, which depends heavily on bin definitions and data distribution.
3. Both approaches can be used in Naive Bayes; Gaussian is more flexible and avoids arbitrary binning but assumes a normal distribution.

Which Method to Use?

Scenario / Data Type	Method
Continuous features approximately Gaussian distributed	Gaussian
Continuous features with unknown/discrete distributions or skewed data	Binning
Small datasets where distribution fitting is unreliable	Binning
When you want to avoid assumptions about data distribution	Binning

Aspect	Gaussian Naive Bayes	Binning (Discretization)
Assumptions	Assumes feature follows Normal distribution	No distributional assumption
Flexibility	Smooth probability density estimation	Depends on chosen bins
Information Loss	Retains continuous nature of data	Some information lost when binning
Parameter Estimation	Estimates mean and variance only	Relies on frequency counts in bins
Handling Outliers	Sensitive to outliers	Outliers may be grouped in bins
Implementation Complexity	Moderate (requires PDF computation)	Simple counting and frequency
Performance	Better if normality holds	Better if actual distribution is skewed or multimodal
Data Size Requirements	Requires sufficient data to estimate parameters	Works well even for smaller datasets
Interpretability	Probabilistic and smooth	Interpretable as class distribution over intervals