# Bayesian Classifiers

- Due No due date
- Points 80
- Questions 8
- Time Limit None
- Allowed Attempts Unlimited

# Instructions

You can have multiple attempt on this quiz to improve your score. Only the highest score will be recorded.

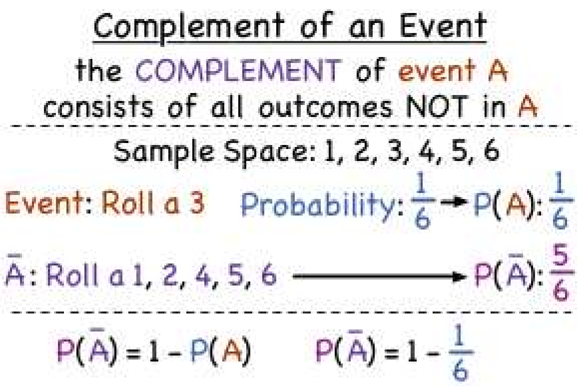
Take the Quiz Again

## Attempt History

| | Attempt | Time | Score |
|---|---|---|---|
| **KEPT** | Attempt 4 | less than 1 minute | 80 out of 80 |
| **LATEST** | Attempt 4 | less than 1 minute | 80 out of 80 |
| | Attempt 3 | 7 minutes | 50 out of 80 |
| | Attempt 2 | 3 minutes | 80 out of 80 |
| | Attempt 1 | 16 minutes | 60 out of 80 |

Score for this attempt: 80 out of 80
Submitted Nov 2 at 7:27pm
This attempt took less than 1 minute.

Question 1
10 / 10 pts




If $P(A) = 0.20$, $P(B|A) = 0.60$ and $P(B|\overline{A}) = 0.25$, then $P(A|B) = ?$

Correct!

0.3750

$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$, everything is given except $P(B)$. Using marginal probability
$P(B) = P(B|A) \times P(A) + P(B|\overline{A}) \times P(\overline{A}) = \mathbf{0.6 \times 0.2 + 0.25 \times (1 - 0.2) = 0.32}$.
Therefore $P(A|B) = \frac{\mathbf{0.6 \times 0.2}}{\mathbf{0.32}} = \mathbf{0.375}$

- 0.7059
- 0.6250
- 0.2941

Question 2

10 / 10 pts

A virus has infected 1.8% of a population. A test detects this virus 95% of the time when it is actually present, but it returns a false positive 3% of the time when the virus is not present.
If a person selected at random from this population tests positive for the virus, what is the probability that this person is actually infected? [Round to the nearest percent.]

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

$$P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A})$$

- 66%
- 63%
- 34%

Correct!

- 37%

If you pay attention to the wording that is used, what you need to calculate is the following conditional probability. We know the test have come up positive so that is the known event.

P ( infected | test is positive ) = ?

Using Bayesian theorem this is equivalent to:

P ( infected | test is positive ) = P ( test is positive | infected ) x P ( infected ) /P ( test is positive )

Among these the two on nominator are already give:

P ( test is positive | infected ) = 0.95

P ( infected ) = 0.018

In order to calculate the denominator we need to use marginal probability.

P ( test is positive ) = P ( test is positive | infected ) x P ( infected ) +P ( test is positive | not infected ) x P ( not infected )

P ( test is positive ) = 0.95 x 0.018 + 0.03 x 0.982 = 0.04656

Finally :

P ( infected | test is positive ) = 0.95 x 0.018 / 0.04656 $\approx$ 0.37

Question 3

10 / 10 pts

$$P(y = K|\vec{x}) = \frac{P(\vec{x}|y = K)P(y = K)}{P(\vec{x})}$$

$P(y = K|\vec{x})$, $P(\vec{x}|y = K)$, $P(y = K)$ and $P(\vec{x})$ are called posterior, likelihood, Prior and Evidence probabilities.

Which statement is correct about the "prior" probability assumptions in Bayesian classifier?

Correct!

- It can be $P(y = K) = \frac{m_K}{m}$, with $m_K$ being the total number of training samples which belong to the category $K$, and $m$ being the total number of training samples.

The prior assumption in Bayesian classifiers is that the probability of belonging to a specific class K is equal to the total number of training data

- None of the above
- It is a probability which must be given prior to collecting the data
- $P(y = K) = \frac{1}{K}$, which means the probability of belonging to a category has to be the same for all $K$ classes.

Question 4

10 / 10 pts

In Bayesian classifiers the best choice of likelihood probability always comes from assuming the Gaussian probability density distribution.

- True

Correct!

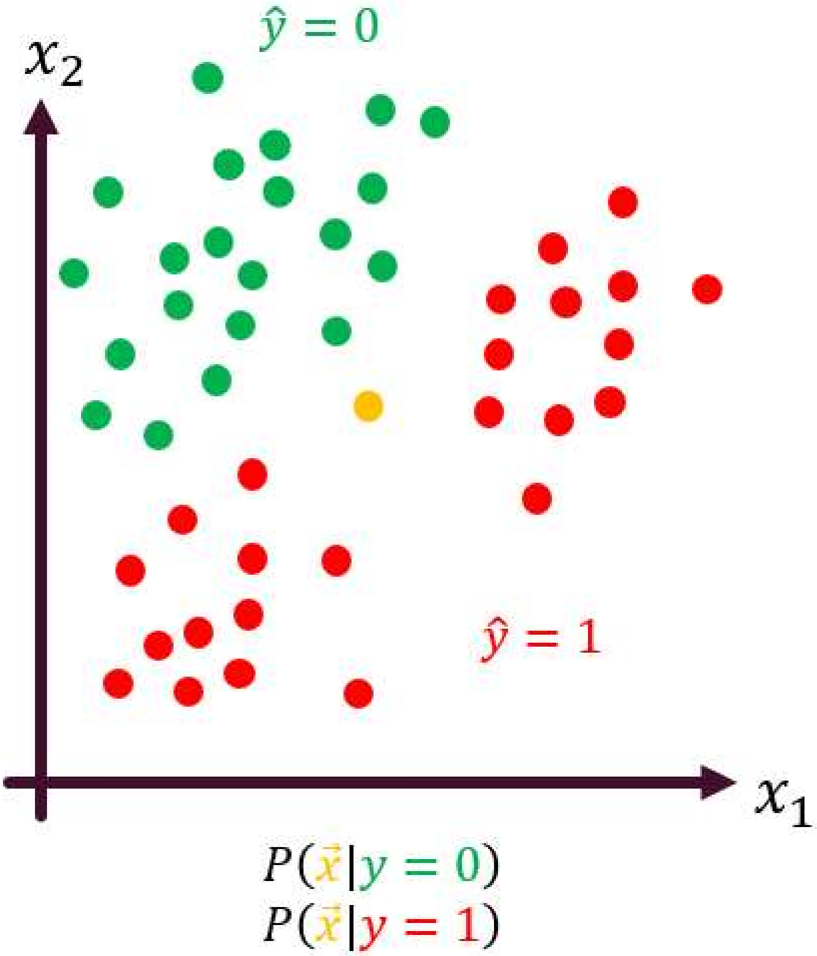
- False

Even though for most problem we do make this assumption but it is not correct to say that Gaussian distribution is "always" the best assumption. Similar to the case of customers visit hours to a restaurant, which we discussed in the class, one should look at the training data and see if Gaussian distribution is a good fit first.

Question 5

10 / 10 pts

In the following categorical example, what is the best assumption for likelihood probabilities of each class?




- [ ] $P(\vec{x}|y=0)$ is best to be modeled by a two dimensional bimodal distribution while $P(\vec{x}|y=1)$ is best to be modeled by a two dimensional Gaussian distribution.
- [ ] Both $P(\vec{x}|y=0)$ and $P(\vec{x}|y=1)$ are best to be modeled by two dimensional Gaussian distribution.
- [ ] Both $P(\vec{x}|y=0)$ and $P(\vec{x}|y=1)$ are best to be modeled by two dimensional bimodal distributions.

Correct!

- [x] $P(\vec{x}|y=0)$ is best to be modeled by a two dimensional Gaussian distribution while $P(\vec{x}|y=1)$ is best to be modeled by a two dimensional bimodal distribution.

That is correct. As it can be seen for category 0 most of the data are concentrated around a single space, but for category 1 the data are concentrated at two different spaces.

Question 6

10 / 10 pts

All the features in Gaussian Naive Bayes are assumed to be uncorrelated.

- [ ] True

Correct!

- [x] False

This is incorrect. The features are assumed to be independent. Independence and uncorrelation are two different things.

Question 7

10 / 10 pts

All the features in Gaussian Naive Bayes are assumed to be independent.

Correct!

- True

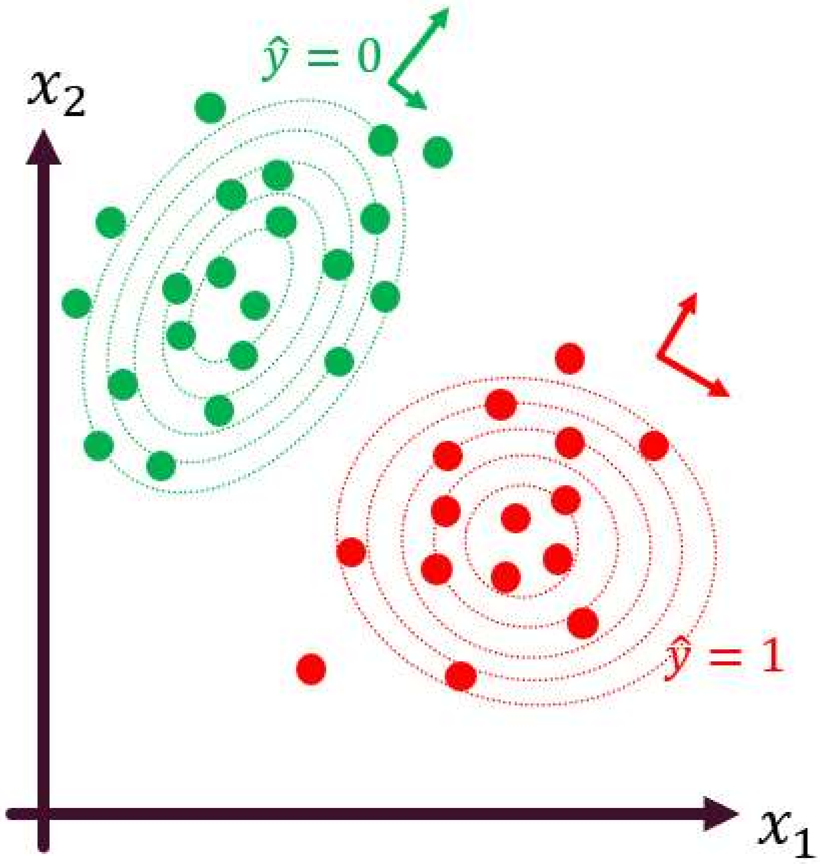
That is correct. Gaussian Naive Bayes classifiers assume all the features are independent.

- False

Question 8

10 / 10 pts

If the underlying probabilistic model of our data is a multivariate normal distribution, which one of the following has the highest chance to fit our training data the best?




- A Linear Discriminant Analysis.
- Gaussian Naive Bayes
- None of the above

Correct!

- A Quadratic Discriminant Analysis.

That is correct. Considering the fact that in general the features are not independent and the fact that different classes might have different covariance matrices the most accurate form of analysis is the QDA. Both LDA and Naive Bayes make extra assumptions which could increase the inaccuracy of the analysis.

Quiz Score: 80 out of 80