Bayes Theorem

Naresh Manwani

Machine Learning Lab, IIIT Hyderabad

Bayes theorem: Describes the probability of an event, based on conditions that might be related to the event.

Bayes theorem: Describes the probability of an event, based on conditions that might be related to the event.

Let us understand using an example !

Let us understand using an example !

- **One fine day Allen felt sick. He went to a doctor.**
- **Doctor suggested a test.**

Let us understand using an example !

- **One fine day Allen felt sick. He went to a doctor.**
- **Doctor suggested a test.**
- **Test result came positive.**

Let us understand using an example !

- One fine day Allen felt sick. He went to a doctor.
- Doctor suggested a test.
- Test result came positive.
- Doctor: "Allen: you have a very rare disease which affects 0.1% of the population in the world."

Let us understand using an example !

- **One fine day Allen felt sick. He went to a doctor.**
- **Doctor suggested a test.**
- **Test result came positive.**

- **Doctor: "Allen: you have a very rare disease which affects 0.1% of the population in the world."**
- **Allen: "How certain is it that I have the disease?"**

Let us understand using an example !

- One fine day Allen felt sick. He went to a doctor.
- Doctor suggested a test.
- Test result came positive.

- Doctor: “Allen: you have a very rare disease which affects 0.1% of the population in the world.”
- Allen: “How certain is it that I have the disease?”
- Doctor: “The test accurately find 99% of the people who have the disease. False alarm rate of the test is 1%.”

Has Allen got disease ?

Has Allen got disease ?

- **What is the chance that Allen actually has the disease?**
- **Is it 99%? (Because that’s the accuracy of the test.)**

Has Allen got disease ?

- What is the chance that Allen actually has the disease?
- Is it 99%? (Because that’s the accuracy of the test.)
- Good news for Allen: that’s not actually correct !

Has Allen got disease ?

- **What is the chance that Allen actually has the disease?**
- **Is it 99%? (Because that's the accuracy of the test.)**
- **Good news for Allen: that's not actually correct !**
- **`Thanks to Bayes Theorem !!!`**

Bayes Theorem

Bayes Theorem

$$P(H|E) = \frac{P(E|H)*P(H)}{P(E|H)*P(H) + P(E|\neg H)*P(\neg H)}$$

Bayes Theorem

$$P(H|E) = \frac{P(E|H)*P(H)}{P(E|H)*P(H) + P(E|\neg H)*P(\neg H)}$$

Probability of hypothesis H being true given that event E has happened

Bayes Theorem

Probability of observing event E given that hypothesis H is true

$$P(H|E) = \frac{P(E|H)*P(H)}{P(E|H)*P(H) + P(E|\neg H)*P(\neg H)}$$

Probability of hypothesis H being true given that event E has happened

Bayes Theorem

Probability of observing event E given that hypothesis H is true

Probability that hypothesis H is true

$$P(H|E) = \frac{P(E|H)*P(H)}{P(E|H)*P(H) + P(E|\neg H)*P(\neg H)}$$

Probability of hypothesis H being true given that event E has happened

Bayes Theorem

Probability of observing event E given that hypothesis H is true

Probability that hypothesis H is true

$$P(H|E) = \frac{P(E|H)*P(H)}{P(E|H)*P(H) + P(E|\neg H)*P(\neg H)}$$

Probability of hypothesis H being true given that event E has happened

Probability of observing event E given that hypothesis H is false

Bayes Theorem

Probability of observing event E given that hypothesis H is true

Probability that hypothesis H is true

$$P(H|E) = \frac{P(E|H)*P(H)}{P(E|H)*P(H) + P(E|\neg H)*P(\neg H)}$$

Probability of hypothesis H being true given that event E has happened

Probability of observing event E given that hypothesis H is false

Probability that hypothesis H is false

Has Allen got disease

Has Allen got disease

- H: Actually have the disease
- E: Test result is positive

Has Allen got disease

- H: Actually have the disease
- E: Test result is positive
- $P(H) = 0.001$

Has Allen got disease

- H: Actually have the disease
- E: Test result is positive
- $P(H) = 0.001$
- $P(\neg H) = 0.999$

Has Allen got disease

- H: Actually have the disease
- E: Test result is positive
- $P(H) = 0.001$
- $P(\neg H) = 0.999$
- $P(E \mid H) = 0.99$

Has Allen got disease

- H: Actually have the disease
- E: Test result is positive
- $P(H) = 0.001$
- $P(\neg H) = 0.999$
- $P(E \mid H) = 0.99$
- $P(E \mid \neg H) = 0.01$ (False alarm rate)

Chance of Allen actually having disease

- Plugging all the numbers, we get

$$P(H|E) = \frac{0.99 * 0.001}{0.99 * 0.001 + 0.01 * 0.999} = 0.09$$

- Only 9% chance that Allen having the disease !

Another Example: Manufacturing Defects in Electric Vehicles

Another Example: Manufacturing Defects in Electric Vehicles

- Suppose that a factory has two machines, Machine A (in China) and Machine B (in India), both producing electric vehicles.
- 40% of their electric vehicles come from Machine A and 60% of their electric vehicles come from Machine B.

Another Example: Manufacturing Defects in Electric Vehicles

- Suppose that a factory has two machines, Machine A (in China) and Machine B (in India), both producing electric vehicles.
- 40% of their electric vehicles come from Machine A and 60% of their electric vehicles come from Machine B.
- 10% of the electric vehicles produced by Machine A are defective.

Another Example: Manufacturing Defects in Electric Vehicles

- Suppose that a factory has two machines, Machine A (in China) and Machine B (in India), both producing electric vehicles.
- 40% of their electric vehicles come from Machine A and 60% of their electric vehicles come from Machine B.
- 10% of the electric vehicles produced by Machine A are defective.
- 5% percent of the electric vehicles from Machine B are defective.

Another Example: Manufacturing Defects in Electric Vehicles

- Suppose that a factory has two machines, Machine A (in China) and Machine B (in India), both producing electric vehicles.
- 40% of their electric vehicles come from Machine A and 60% of their electric vehicles come from Machine B.
- 10% of the electric vehicles produced by Machine A are defective.
- 5% percent of the electric vehicles from Machine B are defective.
- If an electric vehicle is randomly chosen and found defective, what is the probability that it came from machine A?

Probability that Defective EV Came from Machine A

Probability that Defective EV Came from Machine A

- $P(MachineA) = 0.4$
- $P(MachineB) = 0.6$

Probability that Defective EV Came from Machine A

- $P(MachineA) = 0.4$
- $P(MachineB) = 0.6$
- $P(Defective \mid MachineA) = 0.1$

Probability that Defective EV Came from Machine A

- $P(MachineA) = 0.4$
- $P(MachineB) = 0.6$
- $P(Defective \mid MachineA) = 0.1$
- $P(Defective \mid MachineB) = 0.05$

Probability that Defective EV Came from Machine A

- $P(MachineA) = 0.4$
- $P(MachineB) = 0.6$
- $P(Defective \mid MachineA) = 0.1$
- $P(Defective \mid MachineB) = 0.05$
- **Find** $P(MachineA \mid Defective)$

Probability that Defective EV Came from Machine A

- $P(MachineA) = 0.4$
- $P(MachineB) = 0.6$
- $P(Defective \mid MachineA) = 0.1$
- $P(Defective \mid MachineB) = 0.05$
- **Find** $P(MachineA \mid Defective)$

$$P(MachineA \mid Defective) = \frac{P(Defective \mid MachineA)P(MachineA)}{P(Defective \mid MachineA)P(MachineA) + P(Defective \mid MachineB)P(MachineB)}$$

Probability that Defective EV Came from Machine A

- $P(MachineA) = 0.4$
- $P(MachineB) = 0.6$
- $P(Defective \mid MachineA) = 0.1$
- $P(Defective \mid MachineB) = 0.05$
- **Find** $P(MachineA \mid Defective)$

$$P(MachineA \mid Defective) = \frac{P(Defective \mid MachineA)P(MachineA)}{P(Defective \mid MachineA)P(MachineA) + P(Defective \mid MachineB)P(MachineB)}$$

$$= \frac{0.1 * 0.4}{0.1 * 0.4 + 0.05 * 0.6} = \frac{4}{7}$$

Generalized Bayes Theorem

Let $E_1, E_2, \ldots, E_n$ be (pairwise) mutually exclusive events such that $E_1 \cup E_2 \cdots \cup E_n = S$, where S denotes the sample space. Let F be an event such that $P(F) \neq 0$, Then

$$P(E_1 \mid F) = \frac{P(F \mid E_1)P(E_1)}{P(F \mid E_1)P(E_1) + P(F \mid E_2)P(E_2) + \cdots + P(F \mid E_n)P(E_n)}$$

Questions?

Bayes Classifier

Bayes Classifier

Bayes Classifier

- As the name suggests, it is based on Bayes Theorem.
- Lets see …

Bayes Classifier

Bayes Classifier

- Suppose $\mathcal{X}$ be the space from which examples come. $\mathcal{X}$ could be discrete space or continuous.
- For example, in case of text data, $\mathcal{X}$ is bag of words representation of a set of documents.

Bayes Classifier

- Suppose $\mathscr{X}$ be the space from which examples come. $\mathscr{X}$ could be discrete space or continuous.
- For example, in case of text data, $\mathscr{X}$ is bag of words representation of a set of documents.

Bayes Classifier

- Suppose $\mathscr{X}$ be the space from which examples come. $\mathscr{X}$ could be discrete space or continuous.
- For example, in case of text data, $\mathscr{X}$ is bag of words representation of a set of documents.

	the	red	dog	cat	eats	food
1. the red dog	1	1	1	0	0	0
2. cat eats dog	0	0	1	1	1	0
3. dog eats food	0	0	1	0	1	1
4. red cat eats	0	1	0	1	1	0

- $\mathscr{X}$ could be a space of vectors in $\mathbb{R}^d$. Elements of these vectors represent various features of an object. For example, in disease prediction, these feature values can be temperature, blood pressure, oxygen level, heart rate etc.

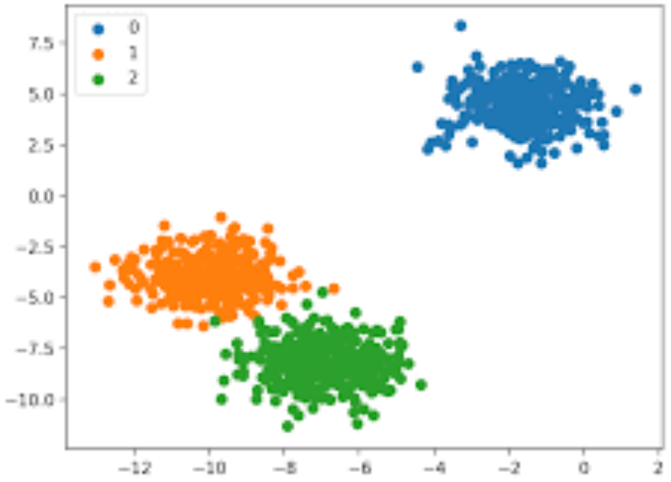
Bayes Classifier

Bayes Classifier

- Suppose there are K categories from which these feature vectors are observed.

Bayes Classifier

- Suppose there are K categories from which these feature vectors are observed.



- The goal is to find a classifier which is inspired by Bayes Theorem.

Bayes Classifier

Bayes Classifier

- Let C denotes the class variable. Thus, $C \in \{1, \ldots, K\}$.
- Let $P(C = i)$ be the prior probability of i^{th} class.

Bayes Classifier

- Let C denotes the class variable. Thus, $C \in \{1, \ldots, K\}$.
- Let $P(C = i)$ be the prior probability of i^{th} class.
- Let $p(\mathbf{x} \,|\, C = i)$ be the class conditional distribution of the i^{th} class.

Bayes Classifier

- Let C denotes the class variable. Thus, $C \in \{1,\ldots,K\}$.
- Let $P(C=i)$ be the prior probability of i^{th} class.
- Let $p(\mathbf{x} \mid C=i)$ be the class conditional distribution of the i^{th} class.
- We first find the posterior distribution of the each class using Bayes Theorem as follows:

Bayes Classifier

- Let C denotes the class variable. Thus, $C \in \{1, \ldots, K\}$.
- Let $P(C = i)$ be the prior probability of i^{th} class.
- Let $p(\mathbf{x} \,|\, C = i)$ be the class conditional distribution of the i^{th} class.
- We first find the posterior distribution of the each class using Bayes Theorem as follows:

$$P(C = i \,|\, \mathbf{x}) = \frac{P(C = i) p(\mathbf{x} \,|\, C = i)}{p(\mathbf{x})}$$

Bayes Classifier

- Let C denotes the class variable. Thus, $C \in \{1,\ldots,K\}$.
- Let $P(C=i)$ be the prior probability of i^{th} class.
- Let $p(\mathbf{x}\,|\,C=i)$ be the class conditional distribution of the i^{th} class.
- We first find the posterior distribution of the each class using Bayes Theorem as follows:

$$P(C=i\,|\,\mathbf{x}) = \frac{P(C=i)p(\mathbf{x}\,|\,C=i)}{p(\mathbf{x})}$$

- Bayes Classifiers predicts class label as follows:

Bayes Classifier

- Let C denotes the class variable. Thus, $C \in \{1, \ldots, K\}$.
- Let $P(C = i)$ be the prior probability of i^{th} class.
- Let $p(\mathbf{x} \mid C = i)$ be the class conditional distribution of the i^{th} class.
- We first find the posterior distribution of the each class using Bayes Theorem as follows:

$$P(C = i \mid \mathbf{x}) = \frac{P(C = i)p(\mathbf{x} \mid C = i)}{p(\mathbf{x})}$$

- Bayes Classifiers predicts class label as follows:

$$\hat{y} = \mathrm{arg}max_{i \in \{1,\ldots,k\}} P(C = i \mid \mathbf{x}) = \mathrm{arg}max_{i \in \{1,\ldots,K\}} P(C = i)p(\mathbf{x} \mid C = i)$$

Quantities to estimate to find Bayes Classifier

- $P(C = i), \forall i$
- $p(\mathbf{x} \mid C = i), \forall i$

Quantities to estimate to find Bayes Classifier

- $P(C = i), \forall i$
- $p(\mathbf{x} \,|\, C = i), \forall i$

- If $\mathbf{x} \in \mathbb{R}^d$, then estimating $p(\mathbf{x} \,|\, C = i)$ means estimating the joint density function of $(x_1, x_2, \ldots, x_d)$ corresponding to i^{th} class.
- This can be a tedious task.

Quantities to estimate to find Bayes Classifier

- $P(C = i), \forall i$
- $p(\mathbf{x} \mid C = i), \forall i$

- If $\mathbf{x} \in \mathbb{R}^d$, then estimating $p(\mathbf{x} \mid C = i)$ means estimating the joint density function of $(x_1, x_2, \ldots, x_d)$ corresponding to i^{th} class.
- This can be a tedious task.
- Naive Bayes makes life easy !

Naive Bayes Assumption

Naive Bayes Assumption

- It assumes that given the class label, features are independent to each other. In other words

$$p(\mathbf{x} \mid C = i) = p(x_1, x_2, \ldots, x_d \mid C = i) = \prod_{k=1}^{d} p(x_k \mid C = i)$$

Naive Bayes Assumption

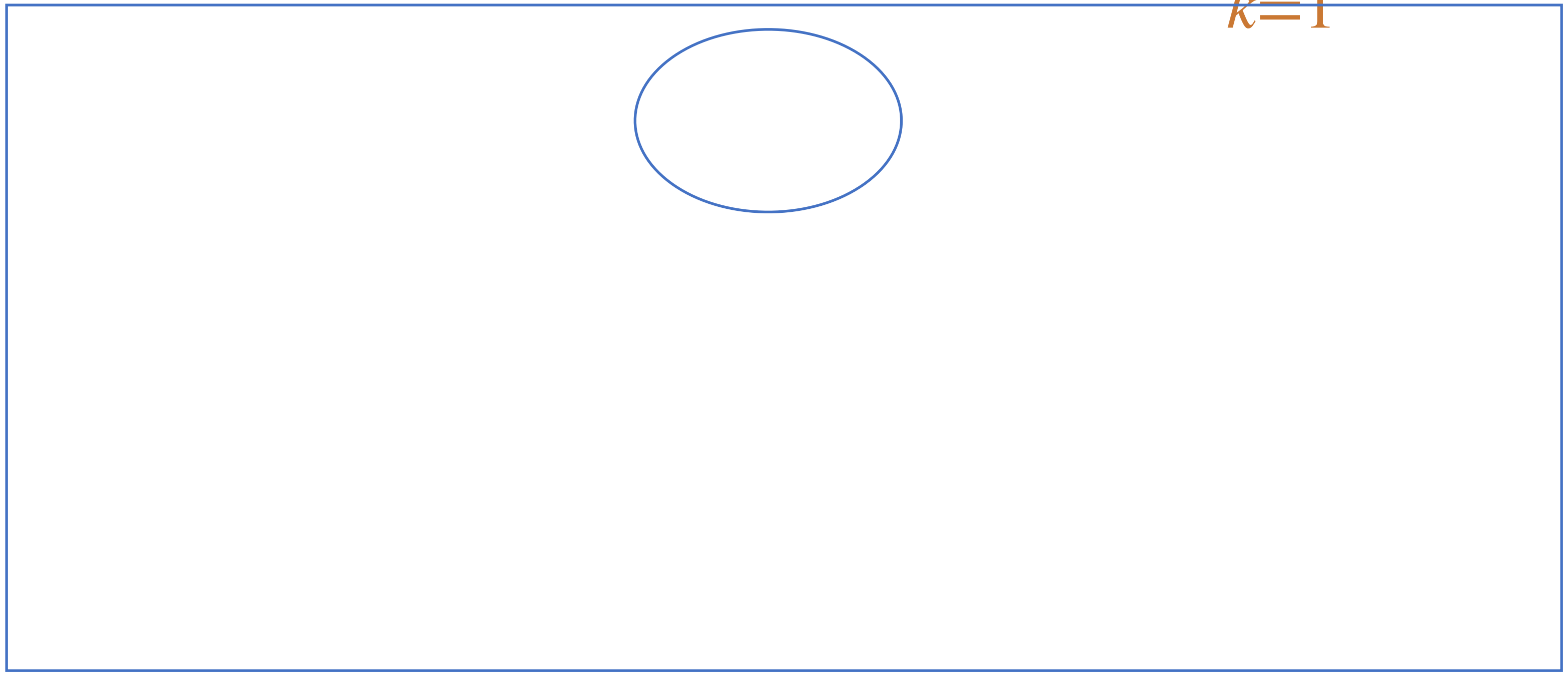
- It assumes that given the class label, features are independent to each other. In other words

$$p(\mathbf{x} \mid C = i) = p(x_1, x_2, \ldots, x_d \mid C = i) = \prod_{k=1}^{d} p(x_k \mid C = i)$$

Naive Bayes Assumption

- It assumes that given the class label, features are independent to each other. In other words

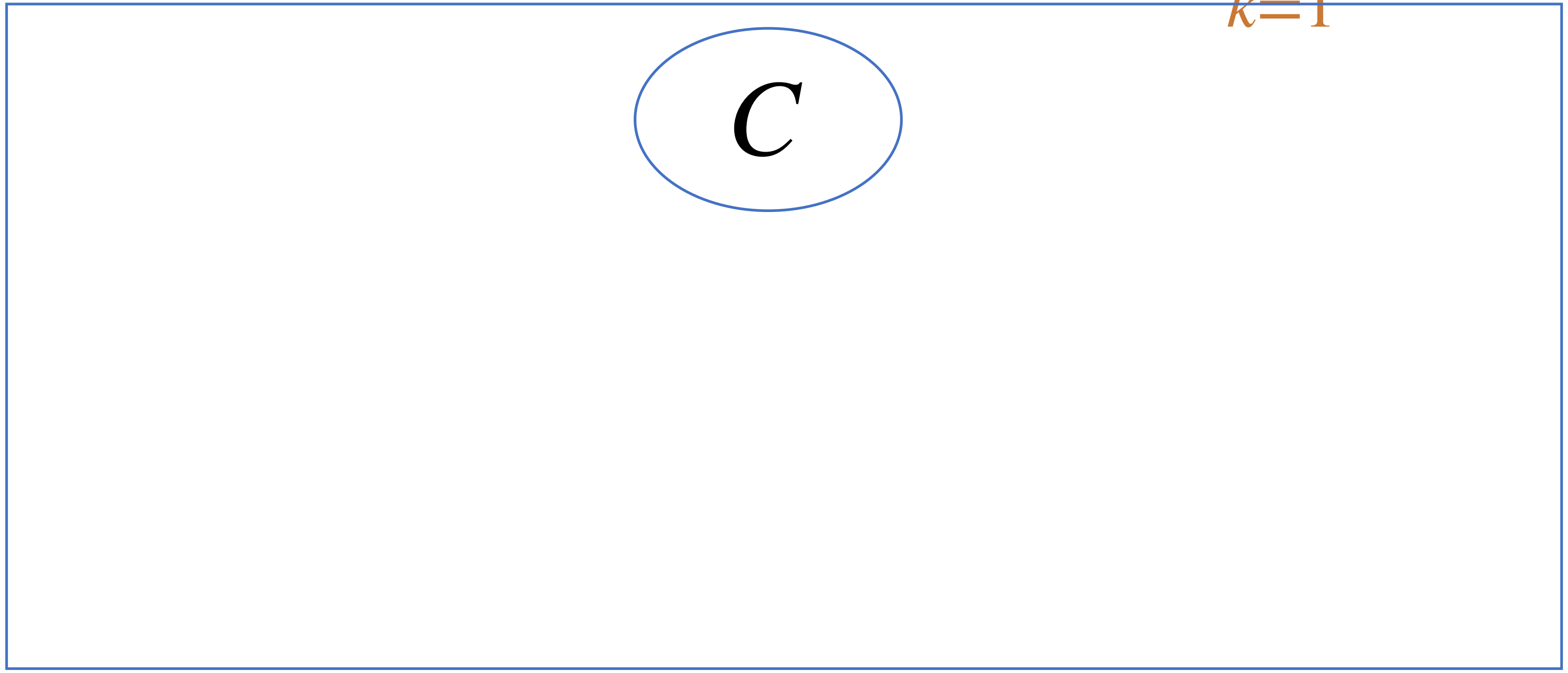
$$p(\mathbf{x} \,|\, C = i) = p(x_1, x_2, \ldots, x_d \,|\, C = i) = \prod_{k=1}^{d} p(x_k \,|\, C = i)$$



Naive Bayes Assumption

- It assumes that given the class label, features are independent to each other. In other words

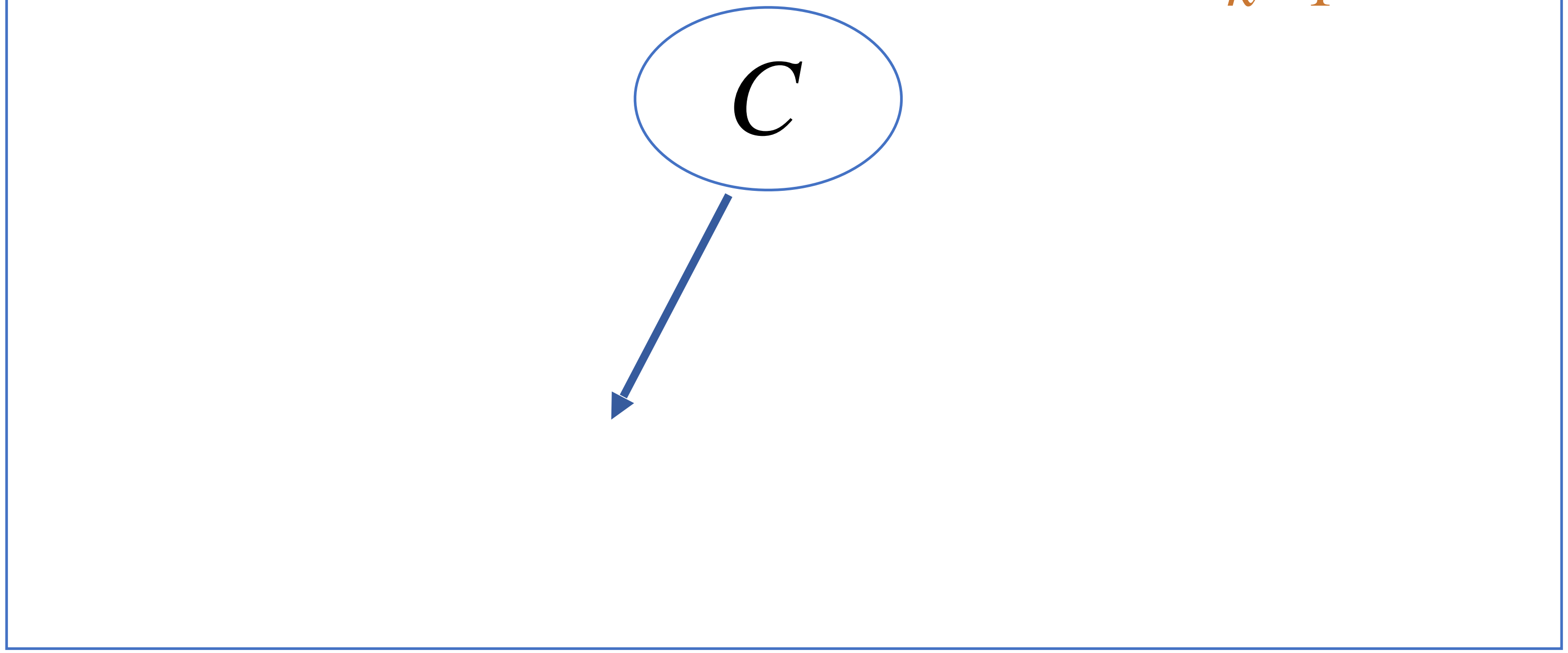
$$p(\mathbf{x} \mid C = i) = p(x_1, x_2, \ldots, x_d \mid C = i) = \prod_{k=1}^{d} p(x_k \mid C = i)$$



Naive Bayes Assumption

- It assumes that given the class label, features are independent to each other. In other words

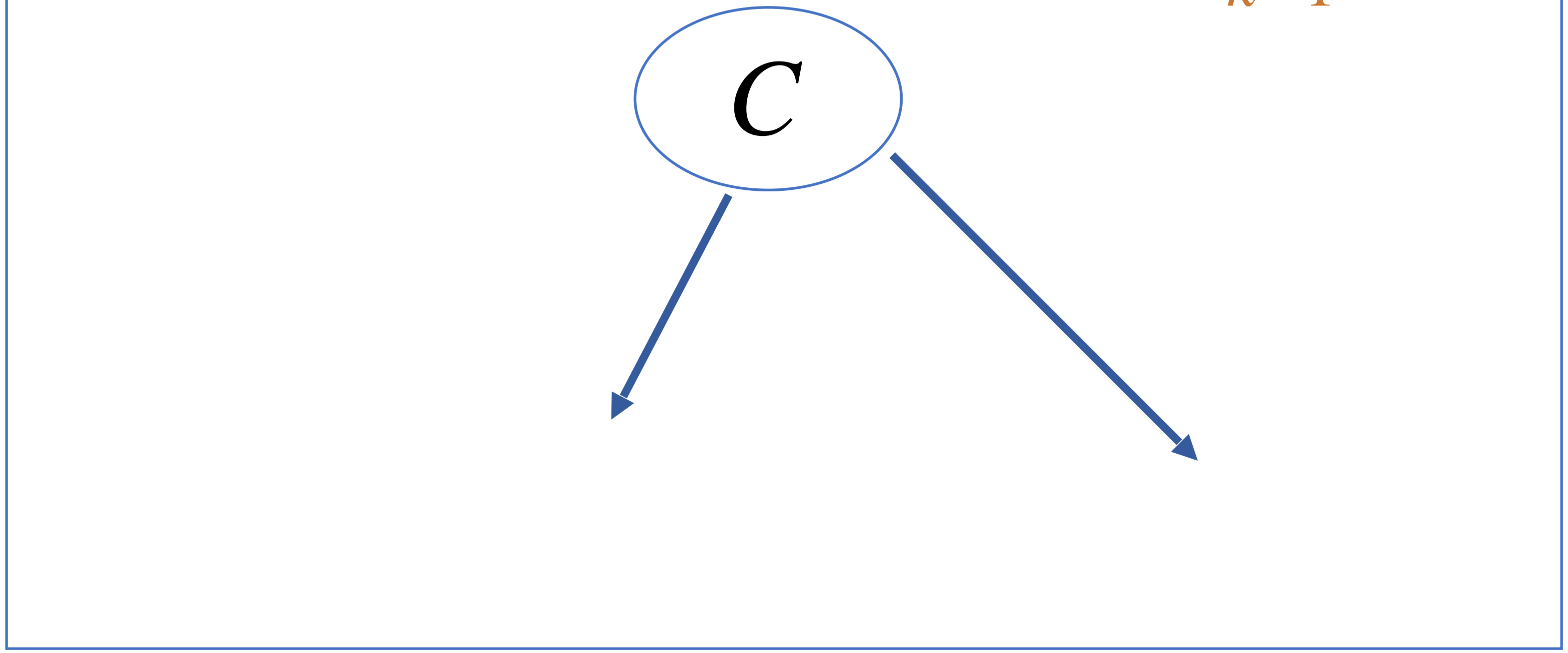
$$p(\mathbf{x} \mid C = i) = p(x_1, x_2, \ldots, x_d \mid C = i) = \prod_{k=1}^{d} p(x_k \mid C = i)$$



Naive Bayes Assumption

- It assumes that given the class label, features are independent to each other. In other words

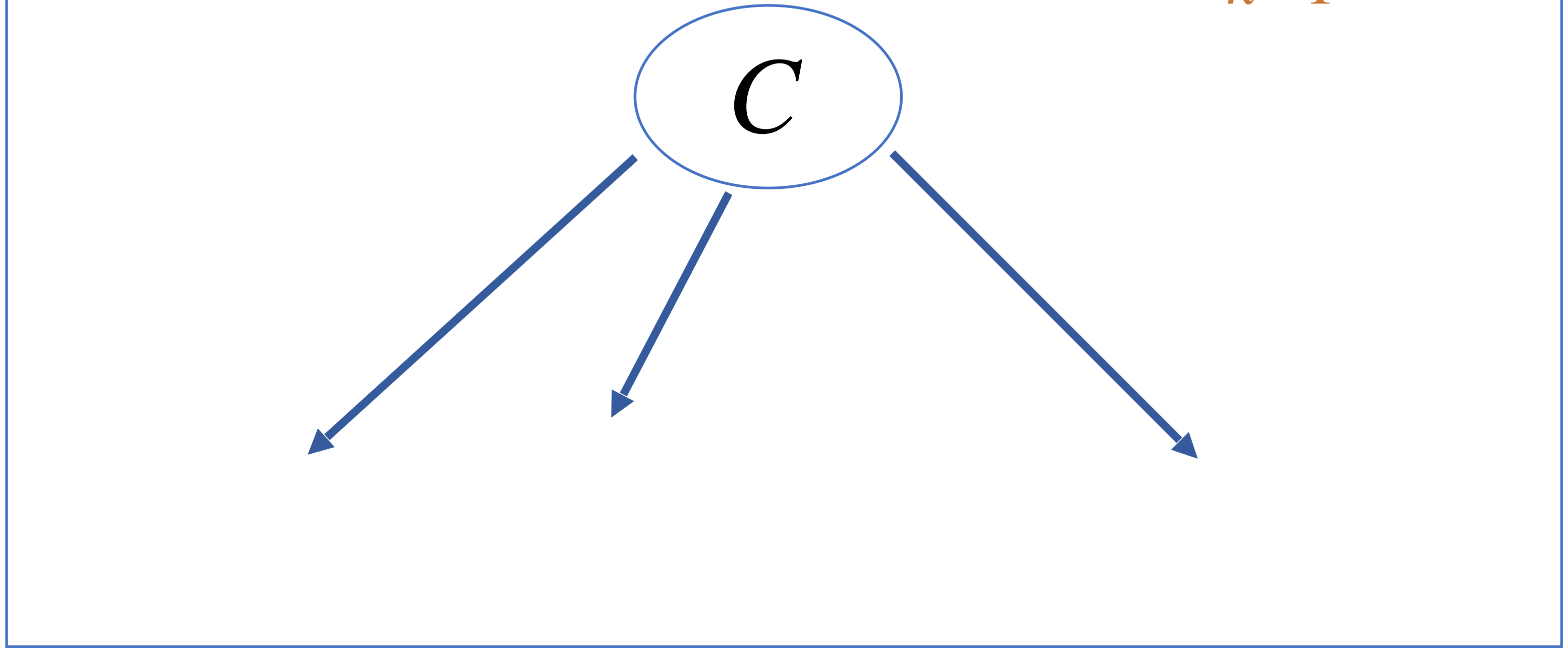
$$p(\mathbf{x} \mid C = i) = p(x_1, x_2, \ldots, x_d \mid C = i) = \prod_{k=1}^{d} p(x_k \mid C = i)$$



Naive Bayes Assumption

- It assumes that given the class label, features are independent to each other. In other words

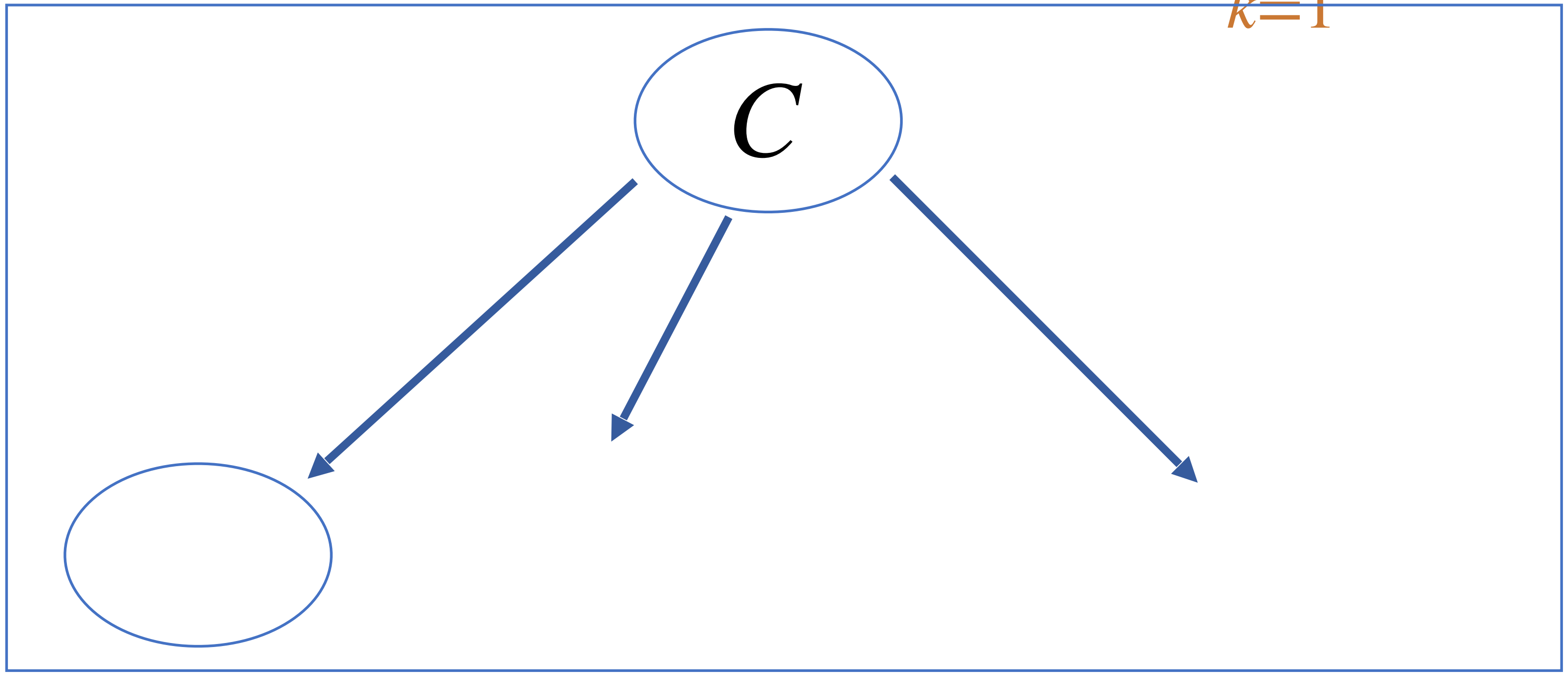
$$p(\mathbf{x} \mid C = i) = p(x_1, x_2, \ldots, x_d \mid C = i) = \prod_{k=1}^{d} p(x_k \mid C = i)$$



Naive Bayes Assumption

- It assumes that given the class label, features are independent to each other. In other words

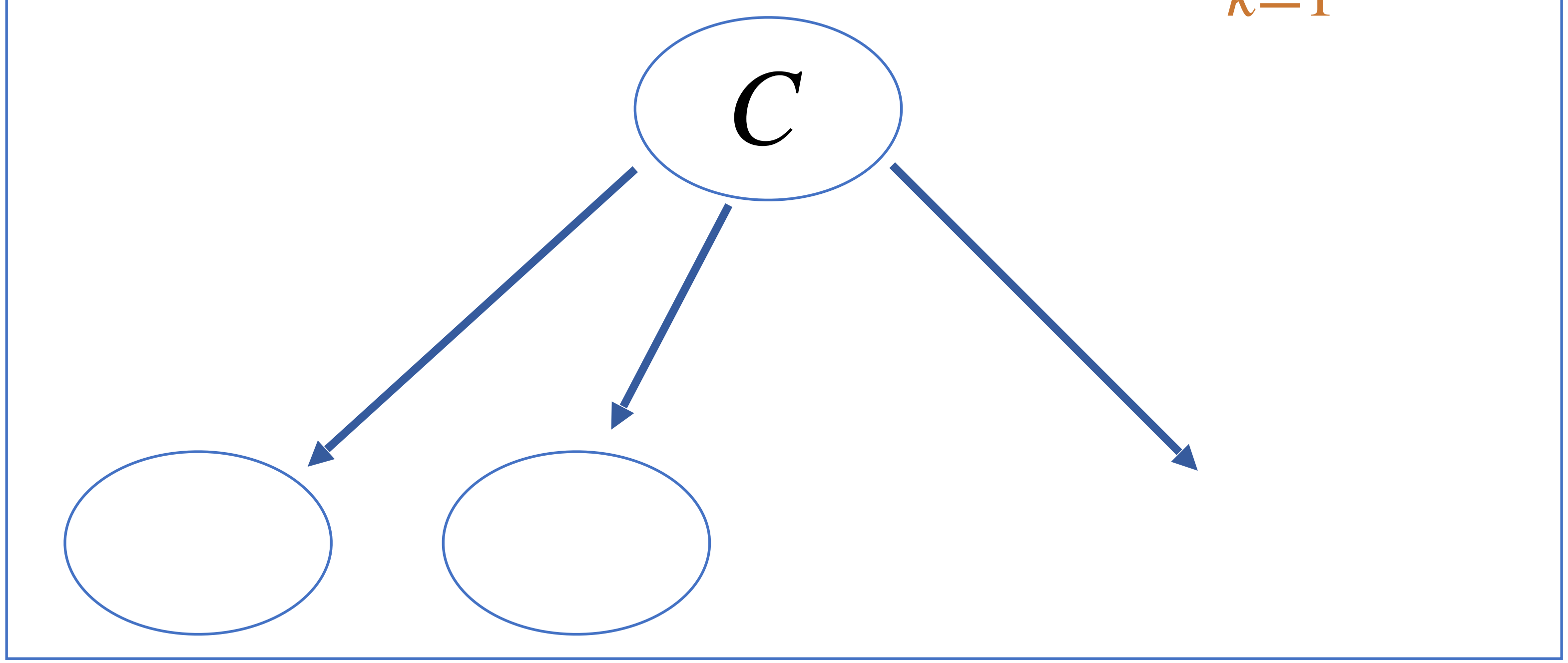
$$p(\mathbf{x} \,|\, C = i) = p(x_1, x_2, \ldots, x_d \,|\, C = i) = \prod_{k=1}^{d} p(x_k \,|\, C = i)$$



Naive Bayes Assumption

- It assumes that given the class label, features are independent to each other. In other words

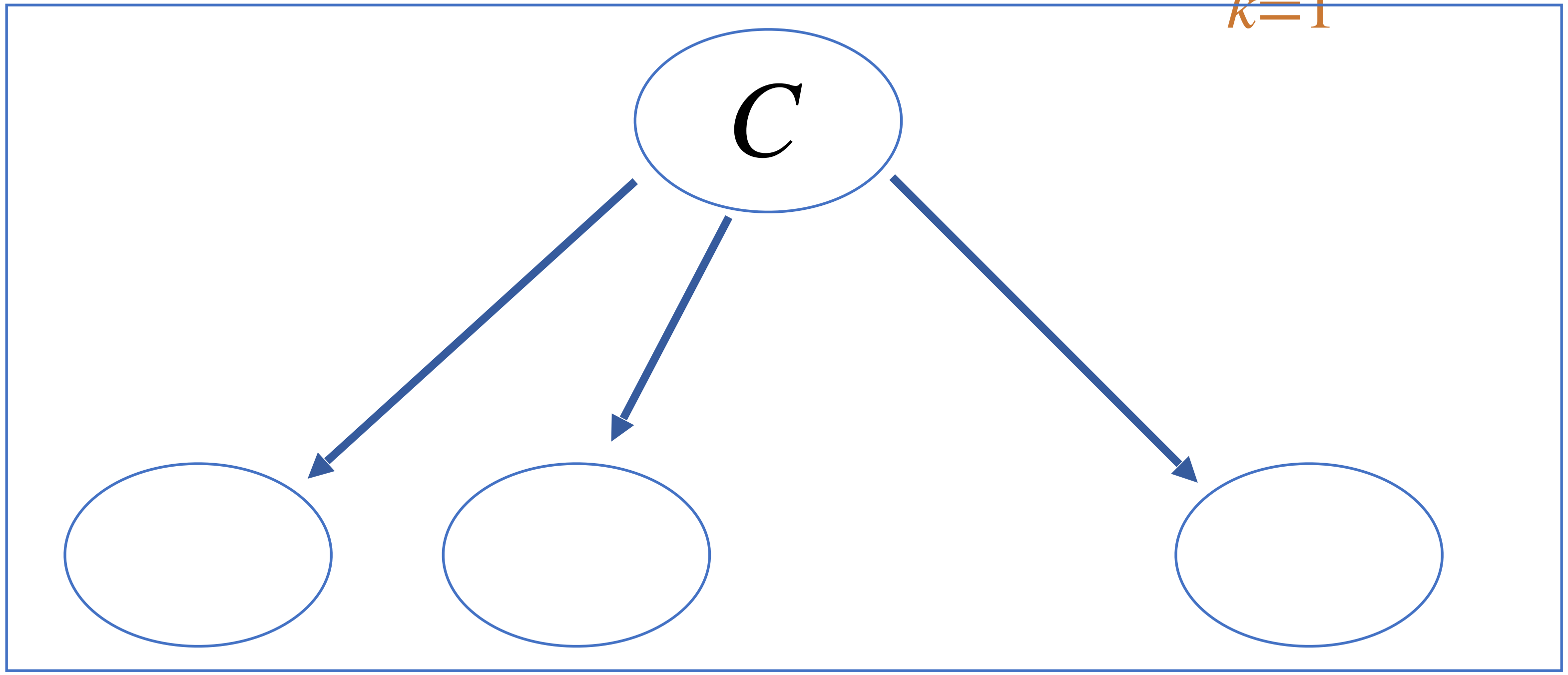
$$p(\mathbf{x} \mid C = i) = p(x_1, x_2, \ldots, x_d \mid C = i) = \prod_{k=1}^{d} p(x_k \mid C = i)$$



Naive Bayes Assumption

- It assumes that given the class label, features are independent to each other. In other words

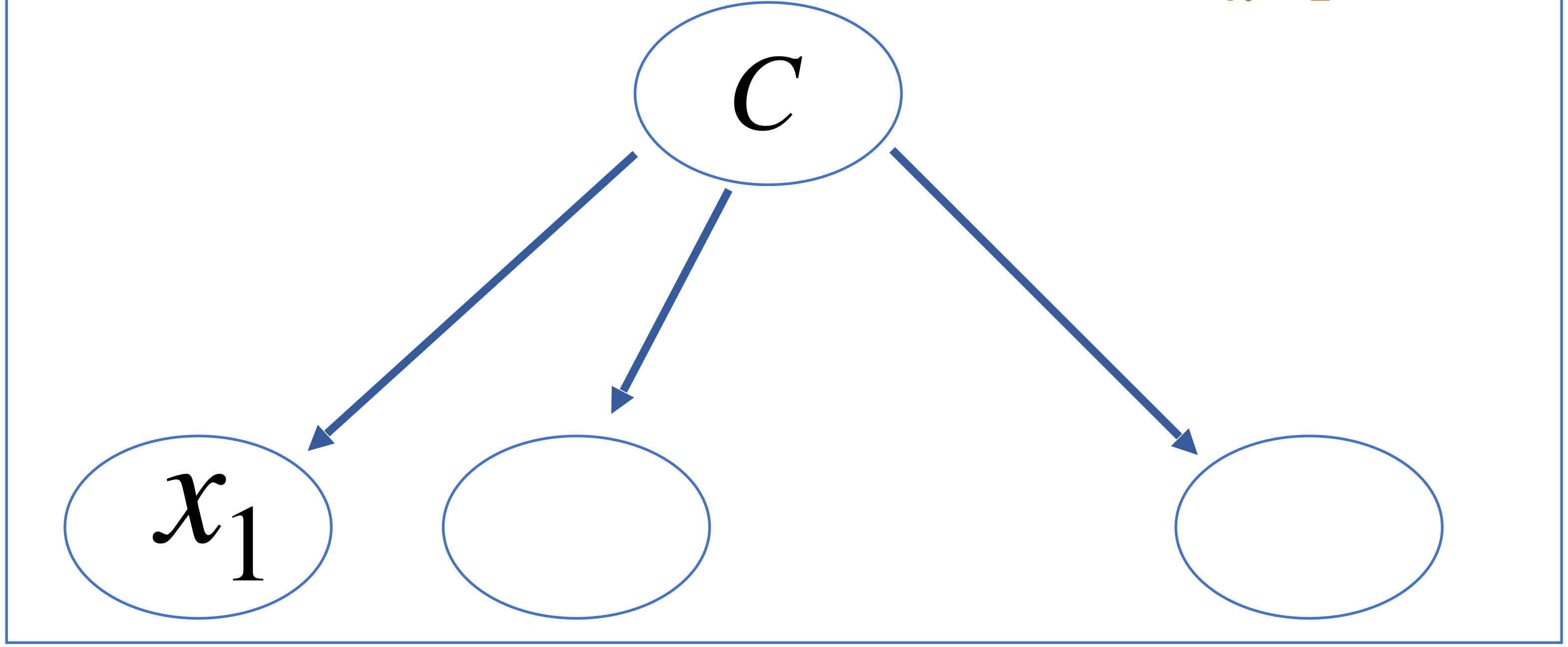
$$p(\mathbf{x} \mid C = i) = p(x_1, x_2, \ldots, x_d \mid C = i) = \prod_{k=1}^{d} p(x_k \mid C = i)$$



Naive Bayes Assumption

- It assumes that given the class label, features are independent to each other. In other words

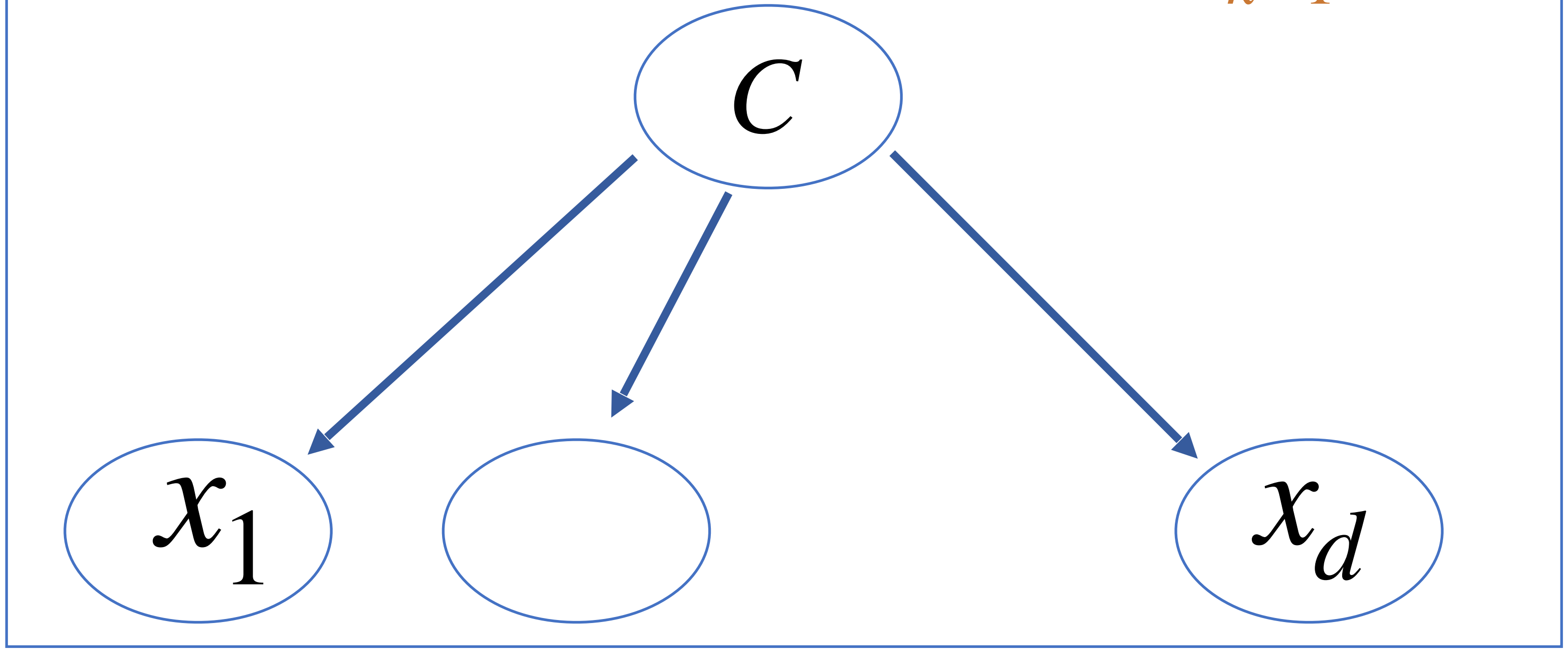
$$p(\mathbf{x} \mid C = i) = p(x_1, x_2, \ldots, x_d \mid C = i) = \prod_{k=1}^{d} p(x_k \mid C = i)$$



Naive Bayes Assumption

- It assumes that given the class label, features are independent to each other. In other words

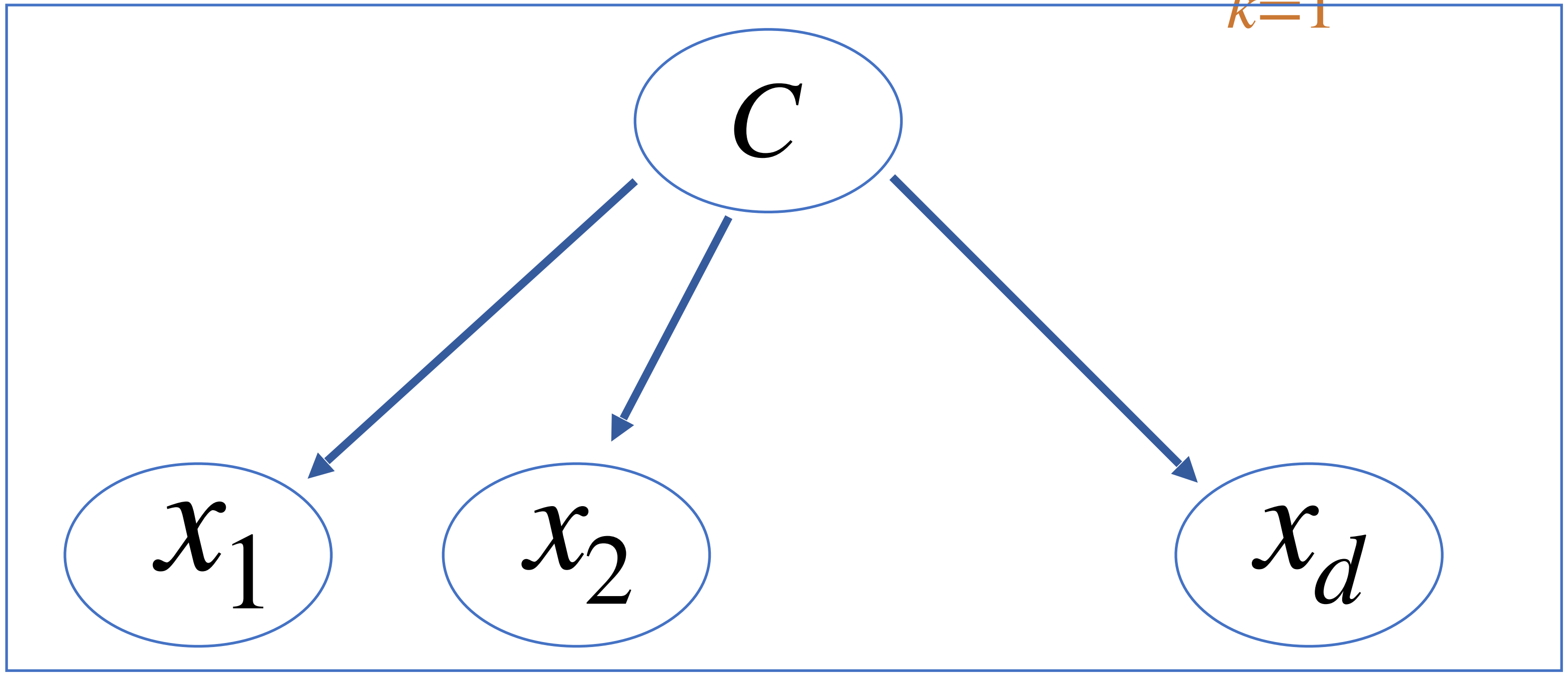
$$p(\mathbf{x} \,|\, C = i) = p(x_1, x_2, \ldots, x_d \,|\, C = i) = \prod_{k=1}^{d} p(x_k \,|\, C = i)$$



Naive Bayes Assumption

- It assumes that given the class label, features are independent to each other. In other words

$$p(\mathbf{x} \mid C = i) = p(x_1, x_2, \ldots, x_d \mid C = i) = \prod_{k=1}^{d} p(x_k \mid C = i)$$



Naive Bayes Classifier

Naive Bayes Classifier

- **Naive Bayes Classifiers predicts class label $\hat{y}$ as follows:**

$$\hat{y} = \mathrm{arg}max_{i\in\{1,\ldots,K\}} P(C=i)p(\mathbf{x}\,|\,C=i)$$

$$= \mathrm{arg}max_{i\in\{1,\ldots,K\}} P(C=i)\prod_{k=1}^{d} p(x_k\,|\,C=i)$$

Quantities to estimate to find Naive Bayes Classifier

Quantities to estimate to find Naive Bayes Classifier

- $P(C = i), \forall i$
- $p(x_k \mid C = i), \forall i, \forall k$

Quantities to estimate to find Naive Bayes Classifier

- $P(C = i), \forall i$
- $p(x_k \mid C = i), \forall i, \forall k$
- **Thus, it only requires the to estimate the distribution of each feature given the class label.**

Quantities to estimate to find Naive Bayes Classifier

- $P(C = i), \forall i$
- $p(x_k \mid C = i), \forall i, \forall k$
- Thus, it only requires the to estimate the distribution of each feature given the class label.
- This not only makes the overall model simple, also reduces the computations required.

Pros:

Pros:

- Simple and fast.

Pros:

- Simple and fast.
- Low computation cost.

Pros:

- Simple and fast.
- Low computation cost.
- Easy and fast to predict class of test data.

Pros:

- Simple and fast.
- Low computation cost.
- Easy and fast to predict class of test data.
- When assumption of independence holds, a Naive Bayes classifier performs better compare to other models like logistic regression and you need less training data.

Cons:

Cons:

- **Zero Frequency Issue:** If categorical variable has a category in test data set, which was not observed in training data set, then model will assign a zero probability and will be unable to make a prediction. This is often known as "Zero Frequency".
- **Independence Assumption:** Another limitation of Naive Bayes is the assumption of independent predictors. In real life, it is almost impossible that we get a set of predictors which are completely independent.

Questions

Examples of Naive Bayes Classifier

Naive Bayes Classifier Example 1:
Predicting whether to play given the weather conditions

Naive Bayes Classifier Example 1:
Predicting whether to play given the weather conditions

Weather Data Record of 14 Days

Day	Outlook	Humidity	Wind	Play
1	Sunny	High	Week	No
2	Sunny	Normal	Strong	Yes
3	Overcast	High	Week	Yes
4	Rain	Normal	Strong	No
5	Rain	Normal	Week	No
6	Sunny	High	Strong	Yes
7	Overcast	Normal	Strong	Yes
8	Sunny	High	Week	Yes
9	Overcast	High	Strong	Yes
10	Rain	Normal	Week	No
11	Overcast	High	Strong	No
12	Rain	High	Week	No
13	Sunny	Normal	Week	No
14	Overcast	High	Strong	Yes

Naive Bayes Classifier

Naive Bayes Classifier

- **Step 1: Find the posterior probabilities as follows:**

Naive Bayes Classifier

- **Step 1: Find the posterior probabilities as follows:**

$$P(Play = Yes \mid Outlook, Humidity, Wind) = \frac{P(Play = Yes)P(Outlook, Humidity, Wind \mid Play = Yes)}{P(Outlook, Humidity, Wind)}$$

$$= \frac{P(Play = Yes)P(Outlook \mid Play = Yes)P(Humidity \mid Play = Yes)P(Wind \mid Play = Yes)}{P(Outlook, Humidity, Wind)}$$

$$P(Play = No \mid Outlook, Humidity, Wind) = \frac{P(Play = No)P(Outlook, Humidity, Wind \mid Play = No)}{P(Outlook, Humidity, Wind)}$$

$$= \frac{P(Play = No)P(Outlook \mid Play = No)P(Humidity \mid Play = No)P(Wind \mid Play = No)}{P(Outlook, Humidity, Wind)}$$

Classification Rule:

Classification Rule:

- **Decide to Play if**

$$P(Play = Yes \mid Outlook, Humidity, Wind) > P(Play = No \mid Outlook, Humidity, Wind)$$

OR

$$P(Play = Yes)P(Outlook \mid Play = Yes)P(Humidity \mid Play = Yes)P(Wind \mid Play = Yes)$$
$$> P(Play = No)P(Outlook \mid Play = No)P(Humidity \mid Play = No)P(Wind \mid Play = No)$$

Classification Rule:

- **Decide to Play if**

$$P(Play = Yes \mid Outlook, Humidity, Wind) > P(Play = No \mid Outlook, Humidity, Wind)$$

OR

$$P(Play = Yes)P(Outlook \mid Play = Yes)P(Humidity \mid Play = Yes)P(Wind \mid Play = Yes)$$
$$> P(Play = No)P(Outlook \mid Play = No)P(Humidity \mid Play = No)P(Wind \mid Play = No)$$

- **Decide Not to Play if**

$$P(Play = No \mid Outlook, Humidity, Wind) > P(Play = Yes \mid Outlook, Humidity, Wind)$$

OR

$$P(Play = No)P(Outlook \mid Play = No)P(Humidity \mid Play = No)P(Wind \mid Play = No)$$
$$> P(Play = Yes)P(Outlook \mid Play = Yes)P(Humidity \mid Play = Yes)P(Wind \mid Play = Yes)$$

Quantities defining the classifier

$$P(Play = Yes) = 0.5$$
$$P(Play = No) = 0.5$$
$$P(Outlook \mid Play = Yes)$$
$$P(Outlook \mid Play = No)$$
$$P(Humidity \mid Play = Yes)$$
$$P(Humidity \mid Play = No)$$
$$P(Wind \mid Play = Yes)$$
$$P(Wind \mid Play = No)$$

Likelihoods of Outlook

Outlook	$P(Outlook \mid Play = Yes)$	$P(Outlook \mid Play = No)$
Sunny	$P(Outlook = Sunny \mid Play = Yes) = \frac{3}{7}$	$P(Outlook = Sunny \mid Play = No) = \frac{2}{7}$
Overcast	$P(Outlook = Overcast \mid Play = Yes) = \frac{4}{7}$	$P(Outlook = Overcast \mid Play = No) = \frac{1}{7}$
Rain	$P(Outlook = Rain \mid Play = Yes) = 0$	$P(Outlook = Rain \mid Play = No) = \frac{4}{7}$

Likelihoods for Humidity

Humidity	$P(Humidity \mid Play = Yes)$	$P(Humidity \mid Play = No)$
High	$P(Humidity = High \mid Play = Yes) = \frac{5}{7}$	$P(Humidity = High \mid Play = No) = \frac{3}{7}$
Normal	$P(Humidity = Normal \mid Play = Yes) = \frac{2}{7}$	$P(Humidity = High \mid Play = No) = \frac{4}{7}$

Likelihoods of Wind

Wind	$P(Wind \mid Play = Yes)$	$P(Wind \mid Play = No)$
Week	$P(Wind = Week \mid Play = Yes) = \frac{2}{7}$	$P(Wind = Week \mid Play = No) = \frac{5}{7}$
Strong	$P(Wind = Strong \mid Play = Yes) = \frac{5}{7}$	$P(Wind = Strong \mid Play = No) = \frac{2}{7}$

Classification for a Test Example

Classification for a Test Example

- Let on a given day, we observe the following: Outlook=Sunny, Humidity=High, Wind=Strong.
- What is the prediction of Naive Bayes Classifier (Play or Not to Play) ?

Classification for a Test Example

- Let on a given day, we observe the following: Outlook=Sunny, Humidity=High, Wind=Strong.
- What is the prediction of Naive Bayes Classifier (Play or Not to Play) ?

- Thus, the classifier predicts to Play

Classification for a Test Example

- **Let on a given day, we observe the following: Outlook=Sunny, Humidity=High, Wind=Strong.**
- **What is the prediction of Naive Bayes Classifier (Play or Not to Play) ?**

$$P(Play = Yes)P(Outlook = Sunny \mid Play = Yes)P(Humidity = High \mid Play = Yes)P(Wind = Strong \mid Play = Yes)$$

$$= 0.5 * \frac{3}{7} * \frac{5}{7} * \frac{5}{7}$$

$$= 0.11$$

- **Thus, the classifier predicts to Play**

Classification for a Test Example

- Let on a given day, we observe the following: Outlook=Sunny, Humidity=High, Wind=Strong.
- What is the prediction of Naive Bayes Classifier (Play or Not to Play) ?

$$P(Play = Yes)P(Outlook = Sunny|Play = Yes)P(Humidity = High|Play = Yes)P(Wind = Strong|Play = Yes)$$
$$= 0.5*\frac{3}{7}*\frac{5}{7}*\frac{5}{7}$$
$$= 0.11$$

$$P(Play = No)P(Outlook = Sunny|Play = No)P(Humidity = High|Play = No)P(Wind = Strong|Play = No)$$
$$= 0.5*\frac{2}{7}*\frac{3}{7}*\frac{2}{7}$$
$$= 0.02$$

- Thus, the classifier predicts to Play

Naive Bayes Classifier Example 2:
Continuous features with Gaussian Distribution

Class Conditional Distribution of Each Feature is Gaussian

Class Conditional Distribution of Each Feature is Gaussian

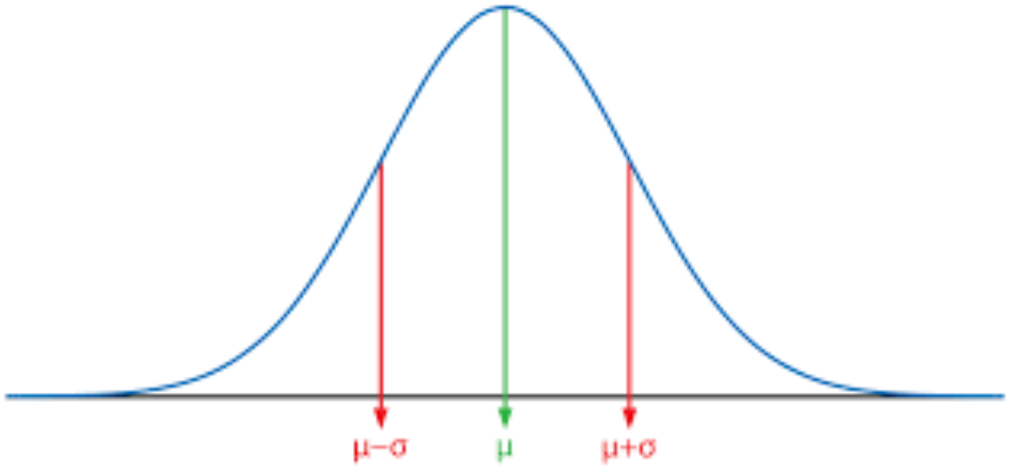
- **Consider binary classification problem.**
- **Let** $p(x_k \mid C = i) = \mathcal{N}(\mu_{i,k}, \sigma_k^2), i = 1,2; k = 1 \ldots d$

Class Conditional Distribution of Each Feature is Gaussian

- **Consider binary classification problem.**
- **Let** $p(x_k \mid C = i) = \mathcal{N}(\mu_{i,k}, \sigma_k^2), i = 1,2; k = 1 \ldots d$
- **Which means, given the class label, each individual feature follows Gaussian distribution**

Class Conditional Distribution of Each Feature is Gaussian

- **Consider binary classification problem.**
- **Let** $p(x_k \mid C = i) = \mathcal{N}(\mu_{i,k}, \sigma_k^2), i = 1,2; k = 1 \ldots d$
- **Which means, given the class label, each individual feature follows Gaussian distribution**



Naive Bayes Classifier

Naive Bayes Classifier

- **For a new example $\mathbf{x}$, the Naive Bayes Classifier predicts class 1 if**

$$P(C=1)\prod_{k=1}^{d} p(x_k \mid C=1) > P(C=2)\prod_{k=1}^{d} p(x_k \mid C=2)$$

$$\Rightarrow \log\left(\prod_{k=1}^{d} \frac{p(x_k \mid C=1)}{p(x_k \mid C=2)}\right) > \log\left(\frac{P(C=2)}{P(C=1)}\right)$$

$$\Rightarrow \sum_{k=1}^{d} \frac{1}{\sigma_k^2}\left[(x_k-\mu_{2,k})^2-(x_k-\mu_{1,k})^2\right] > \log(P(C=2)) - \log(P(C=1))$$

-

$$\Rightarrow \sum_{k=1}^{d} \frac{2\left(\mu_{1,k}-\mu_{2,k}\right)}{\sigma_k^2} x_k > \log(P(C=2)) - \log(P(C=1)) + \sum_{k=1}^{d} \frac{\mu_{1,k}^2-\mu_{2,k}^2}{\sigma_k^2}$$

Naive Bayes Classifier

- For a new example $\mathbf{x}$, the Naive Bayes Classifier predicts class 1 if

$$P(C=1)\prod_{k=1}^{d} p(x_k \mid C=1) > P(C=2)\prod_{k=1}^{d} p(x_k \mid C=2)$$

$$\Rightarrow \log\left(\prod_{k=1}^{d} \frac{p(x_k \mid C=1)}{p(x_k \mid C=2)}\right) > \log\left(\frac{P(C=2)}{P(C=1)}\right)$$

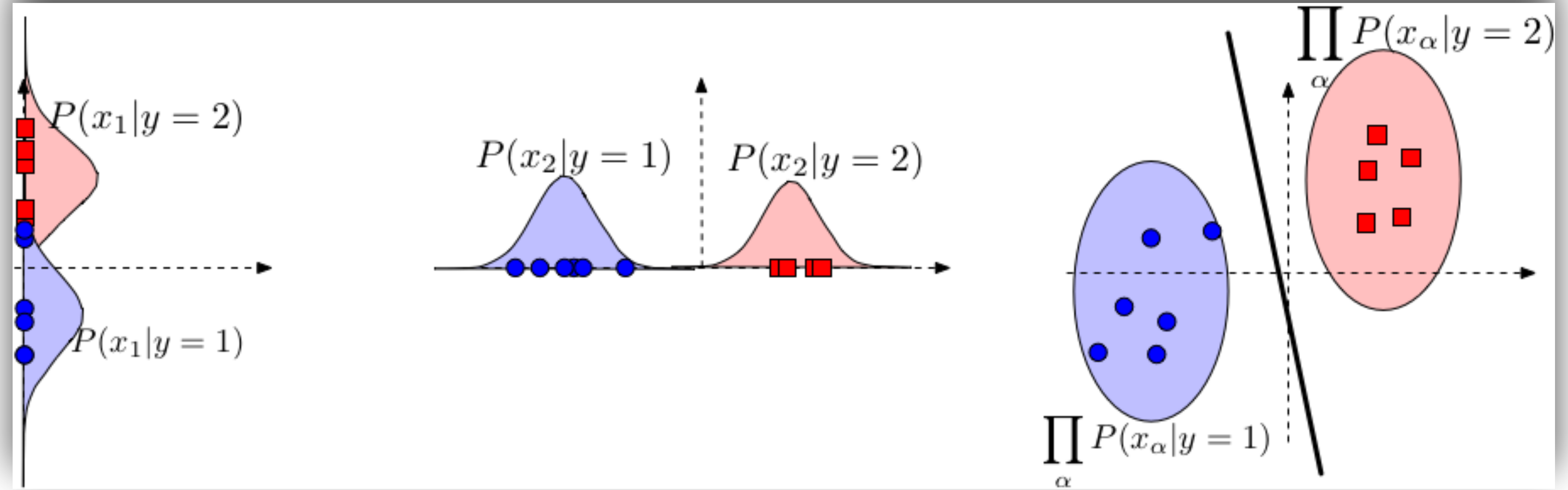
$$\Rightarrow \sum_{k=1}^{d} \frac{1}{\sigma_k^2}\left[(x_k-\mu_{2,k})^2-(x_k-\mu_{1,k})^2\right] > \log(P(C=2)) - \log(P(C=1))$$

-

$$\Rightarrow \sum_{k=1}^{d} \frac{2\left(\mu_{1,k}-\mu_{2,k}\right)}{\sigma_k^2} x_k > \log(P(C=2)) - \log(P(C=1)) + \sum_{k=1}^{d} \frac{\mu_{1,k}^2-\mu_{2,k}^2}{\sigma_k^2}$$

- Naive Bayes Classifier turns out to be Linear

Gaussian Naive Bayes Classifier



Other Applications

Other Applications

- NEWS Classification: Classify the news type (e.g., sports, political, national, international, finance, stock market, cinema, educations etc.) given the news content.
- Spam Mail Or Message Filter

Other Applications

- NEWS Classification: Classify the news type (e.g., sports, political, national, international, finance, stock market, cinema, educations etc.) given the news content.
- Spam Mail Or Message Filter
- Object Detection

Other Applications

- NEWS Classification: Classify the news type (e.g., sports, political, national, international, finance, stock market, cinema, educations etc.) given the news content.
- Spam Mail Or Message Filter
- Object Detection
- Medical Diagnosis

Other Applications

- NEWS Classification: Classify the news type (e.g., sports, political, national, international, finance, stock market, cinema, educations etc.) given the news content.
- Spam Mail Or Message Filter
- Object Detection
- Medical Diagnosis
- Many more ….

scikit-learn Usage

```
>>> from sklearn.datasets import load_iris
>>> from sklearn.model_selection import train_test_split
>>> from sklearn.naive_bayes import GaussianNB
>>> X, y = load_iris(return_X_y=True)
>>> X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.5,
random_state=0)
>>> gnb = GaussianNB()
>>> y_pred = gnb.fit(X_train, y_train).predict(X_test)
```

The documentation contains lot of useful details and explanations

https://scikit-learn.org/stable/modules/naive_bayes.html

Questions?