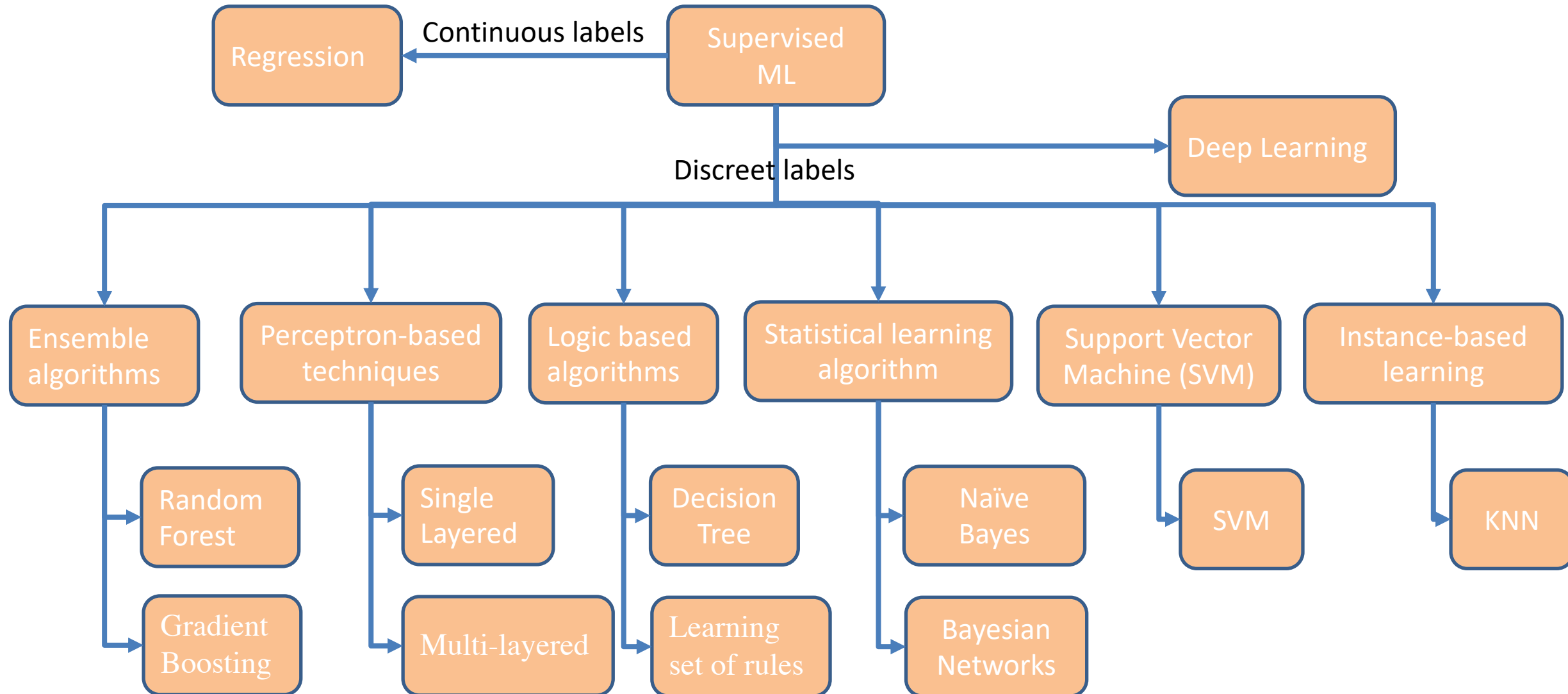


Naïve Bayes Classifier

Today

- Probability Review
- Discrete Naive Bayes classification.
- Continuous Naive Bayes classification.

Types of Supervised ML



Probability

Consider N Events: $e_1, e_2, \dots, e_N$

Conduct Experiment n_T time, and let n_i = The number of times event e_i occurred

Then we can “roughly define” the probability of e_i as $P(e_i) = \frac{n_i}{n_T}$

Example 1 (6 Sided Dice):

A sample space is just the set of all possible results. In simple terms, you have to figure out every possibility for what might happen.

$$e_1 = 1, e_2 = 2, e_3 = 3, e_4 = 4, e_5 = 5, e_6 = 6$$

Theory we know that $P(e_i) = \frac{1}{6}$

Random Variable

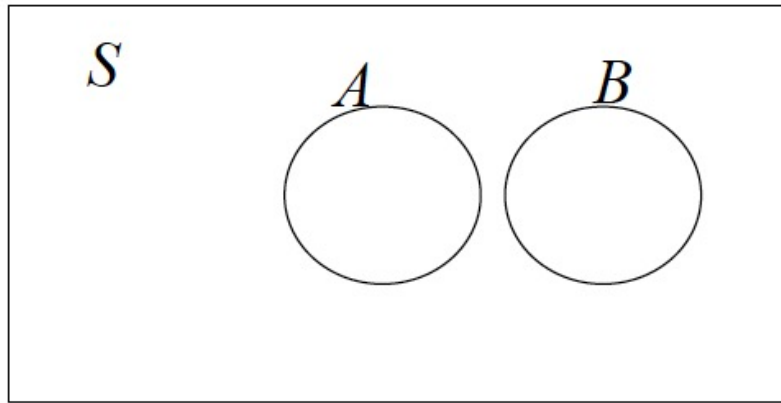
Definition: A variable that can take on several values, each value having a probability of occurrence.

There are two types of random variables:

- Discrete. Take on a countable number of values.
- Continuous. Take on a range of values.

The Sample Space

The space of all possible events of a given process or situation is called the sample space S . $S = \{e_1 = 1, e_2 = 2, e_3 = 3, e_4 = 4, e_5 = 5, e_6 = 6\}$



1. $A = \{1, 2\}$ $B = \{3\}$

$$P(A \cup B) = P(\omega_i \leq 2) + P(\omega_i = 3) \\ = \frac{2}{6} + \frac{1}{6} = \frac{1}{2} \quad \text{as before}$$

2. $A = \{1, 2\}$ $B = \{3\}$

$$P(A \cup B) \neq P(A) + P(B)$$

$= P(A)$

$$P(A) \neq P(A) + \frac{1}{6}$$

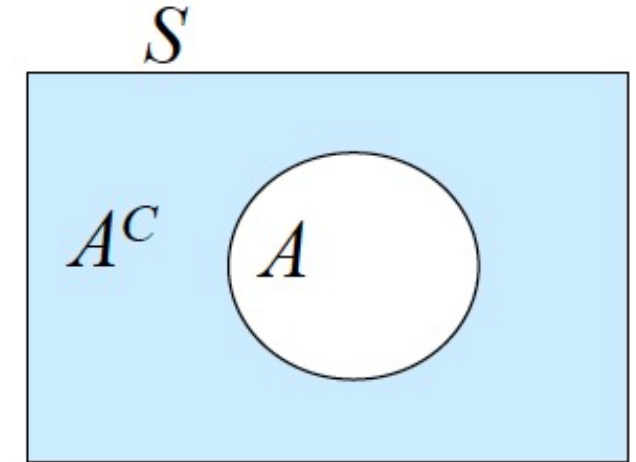
Rules of probability

$$P1: P(A^c \cup A) = 1 \quad (\text{Because } S = A \cup A^c)$$

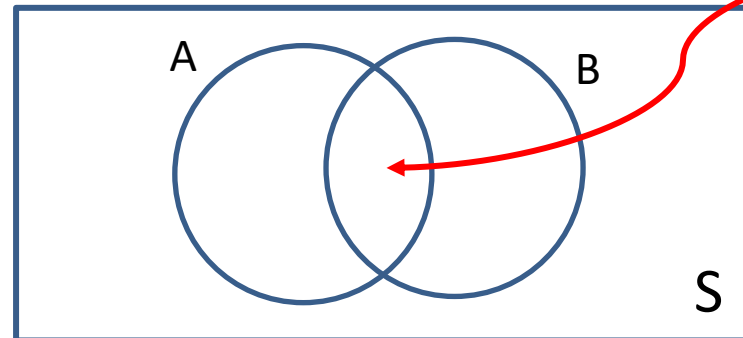
$$P2: P(A^c \cup A) = P(A^c) + P(A)$$

Follows from A3 because $A \cap A^c = \emptyset$

P1 & P2 together give $P(A^c) = 1 - P(A)$



$$P3: \text{If } A \cap B \neq \emptyset, \text{ then } P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



$A \cap B$ is counted twice

Joint Probability

Consider two separate “experiments”: The probability that...

The 1st experiment had outcome A, **AND** the 2nd experiment had outcome B is denoted as $P(A,B)$

Often A& B come from a single experiment having multiple observations

Suppose we collected a data of people aged between 18 and 45, the data includes age, Height & Weight

$P(H > 1.70 \text{ M} , W > 75 \text{ Kg}) =$ The probability the selected person is taller than 1.70
AND weighs more than 75 Kg = $\frac{\text{The \#of observations where}(H=1.7 \text{ and } W=75)}{\text{The total \#of observations}}$

Conditional Probability & Independence

Consider two separate observations. Given that you know what was observed for one of the outcomes, what is the probability that you will get the other outcome?

$P(A|B)$ = probability that you observe A given that B has occurred

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

If A and B are independent to each other then the knowing B won't change the probability of observing A:

$$P(A|B) = P(A)$$

$$P(A, B) = P(A|B)P(B) = P(A)P(B)$$

Example

An urn contains 6 red marbles and 4 black marbles. Two marbles are drawn *without replacement* from the urn. What is the probability that both of the marbles are black?

Solution:

Let A = the event that the first marble is black; and let B = the event that the second marble is black. We know the following:

In the beginning, there are 10 marbles in the urn, 4 of which are black. Therefore,
 $P(A) = 4/10$.

After the first selection, there are 9 marbles in the urn, 3 of which are black. Therefore,
 $P(B|A) = 3/9$.

Therefore, based on the rule of multiplication:

$$P(A, B) = P(A) P(B|A)$$

$$P(A, B) = (4/10) * (3/9) = 12/90 = 2/15$$

Example

A student goes to the library. The probability that she checks out (a) a work of fiction is 0.40, (b) a work of non-fiction is 0.30, and (c) both fiction and non-fiction is 0.20. What is the probability that the student checks out a work of fiction, non-fiction, or both?

Solution:

Let F = the event that the student checks out fiction; and let N = the event that the student checks out non-fiction. Then, based on the rule of addition:

$$P(F \cup N) = P(F) + P(N) - P(F \cap N)$$

$$P(F \cup N) = 0.40 + 0.30 - 0.20 = 0.50$$

Test Your Understanding

Problem 1

An urn contains 6 red marbles and 4 black marbles. Two marbles are drawn *with replacement* from the urn. What is the probability that both of the marbles are black?

- (A) 0.16
- (B) 0.32
- (C) 0.36
- (D) 0.40
- (E) 0.60

Solution

The correct answer is A. Let A = the event that the first marble is black; and let B = the event that the second marble is black. We know the following:

In the beginning, there are 10 marbles in the urn, 4 of which are black. Therefore, $P(A) = 4/10$.

After the first selection, we replace the selected marble; so there are still 10 marbles in the urn, 4 of which are black. Therefore, $P(B|A) = 4/10$.

Therefore, based on the rule of multiplication:

$$P(A, B) = P(A) P(B|A)$$

$$P(A, B) = (4/10) * (4/10) = 16/100 = 0.16$$

Problem 2

A card is drawn randomly from a deck of ordinary playing cards. You win \$10 if the card is a spade or an ace. What is the probability that you will win the game?

(A) $1/13$

(B) $13/52$

(C) $4/13$

(D) $17/52$

(E) None of the above.

Solution

The correct answer is C. Let S = the event that the card is a spade; and let A = the event that the card is an ace. We know the following:

There are 52 cards in the deck.

There are 13 spades, so $P(S) = 13/52$.

There are 4 aces, so $P(A) = 4/52$.

There is 1 ace that is also a spade, so $P(S \cap A) = 1/52$.

Therefore, based on the rule of addition:

$$P(S \cup A) = P(S) + P(A) - P(S \cap A)$$

$$P(S \cup A) = 13/52 + 4/52 - 1/52 = 16/52 = 4/13$$

Bayesian decision theory

Bayesian decision theory is a fundamental statistical approach to the problem of pattern classification. This approach is based on quantifying the trade-offs between various classification decisions using probability and the costs that accompany such decisions.

Illustrative example

Suppose that an observer watching fish arrive along the conveyor belt finds it hard to predict what type will emerge next and that the sequence of types of fish appears to be random.

ω denote the state state of of nature

$\omega = \omega_1$ for sea bass

$\omega = \omega_2$ for salmon

Let us assume that there is some a priori probability calculated based on previously collected data

$P(\omega_1)$ = The priori probability that the fish is sea bass

$P(\omega_2)$ = The priori probability that the fish is salmon

$P(\omega_1) + P(\omega_2) = 1$

These prior probabilities reflect our prior knowledge of how likely we are to get a sea bass or salmon before the fish actually appears.

How to make a decision about the type of fish that will appear next without being allowed to see it?

One answer can be: Decide ω_1 if $P(\omega_1) > P(\omega_2)$; otherwise decide ω_2

This rule makes sense if we are to judge just one fish. But in the case of many fish, we will be making the same prediction although we know that both fish can appear

Suppose we are given with the data a feature x that describes the fish lightness readings.

$p(x | \omega_1)$ = the probability of x given the fish is sea bass.

$p(x | \omega_2)$ = the probability of x given the fish is sea bass.

The information we need to calculate this conditional probability is available in the data

Suppose we have calculated the probabilities ($P(\omega_1)$, $P(\omega_2)$, $p(x|\omega_1)$, and $p(x|\omega_2)$) based on the data we have.

How can we predict the next fish if we know the value is x (lightness of the fish)?

$$p(\omega_j, x) = P(\omega_j | x)p(x) = p(x|\omega_j)P(\omega_j)$$

Rearranging these leads us to the answer to our question, which is called *Bayes' Formula*

$$p(\omega_j | x) = \frac{p(x|\omega_j)p(\omega_j)}{p(x)}$$

$$posterior = \frac{likelihood \times prior}{evidence}$$

$$p(\omega_j|x) = \frac{p(x|\omega_j)p(\omega_j)}{p(x)}$$

$$posterior = \frac{likelihood \times prior}{evidence}$$

Where :

$P(\omega_j)$ to the a posteriori probability.

$p(x|\omega_j)$ the likelihood of ω_j with respect to x

$P(x)$ is the prior probability of *predictor*.

$P(\omega_j | x)$ posterior — the probability of the state of nature being ω_j given that feature value x has been measured.

In the case of the data with more than one features $(x_1, x_2, \dots, x_N)$ we have calculate $P(\omega_j | x_1, x_2, \dots, x_N)$ for all the features.

$$P(\omega, x_1, x_2, \dots, x_N) = P(x_1, x_2, \dots, x_N, \omega)$$

$$P(\omega, x_1, x_2, \dots, x_N) = P(x_1 | x_2, \dots, x_N, \omega) P(x_2 | x_3, \dots, x_N, \omega) \dots P(x_N | \omega) P(\omega)$$

Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

$$P(\omega, x_1, x_2, \dots, x_N) = P(\omega) \prod_{i=1}^N P(x_i | \omega)$$

Constructing a classifier from the probability model

The discussion so far has derived the independent feature model, that is, the naïve Bayes probability model.

$$y = \operatorname{argmax}_{k=\{1,2,\dots,k\}} P(\omega_k) \prod_{i=1}^N P(x_i, |\omega_k)$$

Naïve Bayes

- Naïve Bayes Algorithm (for discrete input attributes)

- **Learning Phase:** Given a training set S ,

For each target value of c_i ($c_i = c_1, \dots, c_L$)

$\hat{P}(C = c_i) \leftarrow$ estimate $P(C = c_i)$ with examples in S ;

For every attribute value a_{jk} of each attribute x_j ($j = 1, \dots, n; k = 1, \dots, N_j$)

$\hat{P}(X_j = a_{jk} | C = c_i) \leftarrow$ estimate $P(X_j = a_{jk} | C = c_i)$ with examples in S ;

Output: conditional probability tables; for x_j , $N_j \times L$ elements

- **Test Phase:** Given an unknown instance $\mathbf{X}' = (a'_1, \dots, a'_n)$,

Look up tables to assign the label c^* to $\mathbf{X}'$ if

$$[\hat{P}(a'_1 | c^*) \cdots \hat{P}(a'_n | c^*)] \hat{P}(c^*) > [\hat{P}(a'_1 | c) \cdots \hat{P}(a'_n | c)] \hat{P}(c), \quad c \neq c^*, c = c_1, \dots, c_L$$

Do I play Tennis today? An example.

PlayTennis: training examples

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

- Learning Phase

Outlook	Play=Yes	Play=No
<i>Sunny</i>	2/9	3/5
<i>Overcast</i>	4/9	0/5
<i>Rain</i>	3/9	2/5

Temperature	Play=Yes	Play=No
<i>Hot</i>	2/9	2/5
<i>Mild</i>	4/9	2/5
<i>Cool</i>	3/9	1/5

Humidity	Play=Yes	Play=No
<i>High</i>	3/9	4/5
<i>Normal</i>	6/9	1/5

Wind	Play=Yes	Play=No
<i>Strong</i>	3/9	3/5
<i>Weak</i>	6/9	2/5

$$P(\text{Play=Yes}) = 9/14 \quad P(\text{Play=No}) = 5/14$$

- Test Phase

- Given a new instance,

$\mathbf{x}' = (\text{Outlook}=\text{Sunny}, \text{Temperature}=\text{Cool}, \text{Humidity}=\text{High}, \text{Wind}=\text{Strong})$

- Look up tables

$$P(\text{Outlook}=\text{Sunny} \mid \text{Play}=\text{Yes}) = 2/9$$

$$P(\text{Outlook}=\text{Sunny} \mid \text{Play}=\text{No}) = 3/5$$

$$P(\text{Temperature}=\text{Cool} \mid \text{Play}=\text{Yes}) = 3/9$$

$$P(\text{Temperature}=\text{Cool} \mid \text{Play}=\text{No}) = 1/5$$

$$P(\text{Humidity}=\text{High} \mid \text{Play}=\text{Yes}) = 3/9$$

$$P(\text{Humidity}=\text{High} \mid \text{Play}=\text{No}) = 4/5$$

$$P(\text{Wind}=\text{Strong} \mid \text{Play}=\text{Yes}) = 3/9$$

$$P(\text{Wind}=\text{Strong} \mid \text{Play}=\text{No}) = 3/5$$

$$P(\text{Play}=\text{Yes}) = 9/14$$

$$P(\text{Play}=\text{No}) = 5/14$$

- MAP rule

$$P(\text{Yes} \mid \mathbf{x}'): [P(\text{Sunny} \mid \text{Yes})P(\text{Cool} \mid \text{Yes})P(\text{High} \mid \text{Yes})P(\text{Strong} \mid \text{Yes})]P(\text{Play}=\text{Yes}) = 0.0053$$

$$P(\text{No} \mid \mathbf{x}'): [P(\text{Sunny} \mid \text{No})P(\text{Cool} \mid \text{No})P(\text{High} \mid \text{No})P(\text{Strong} \mid \text{No})]P(\text{Play}=\text{No}) = 0.0206$$

Given the fact $P(\text{Yes} \mid \mathbf{x}') < P(\text{No} \mid \mathbf{x}')$, we label $\mathbf{x}'$ to be “No”.

Continuous Naïve Bayes

Continuous Naïve Bayes

We've now covered Discrete Naïve Bayes and tried an example.

Discrete Naïve Bayes dealt with conditional probabilities of discrete class memberships, for example `Character_Type={Ogre, Human, Orc, etc}` or `Speed={Fast,Slow}`

What happens when we deal with real, continuous numbers? For example Temperature?

We could convert a temperature to a discrete class, eg if $(12 \leq \text{temp} \leq 16)$ then we set this instance to be a member of class 'cool' and we might similarly have classes for 'hot' and 'just_right.'

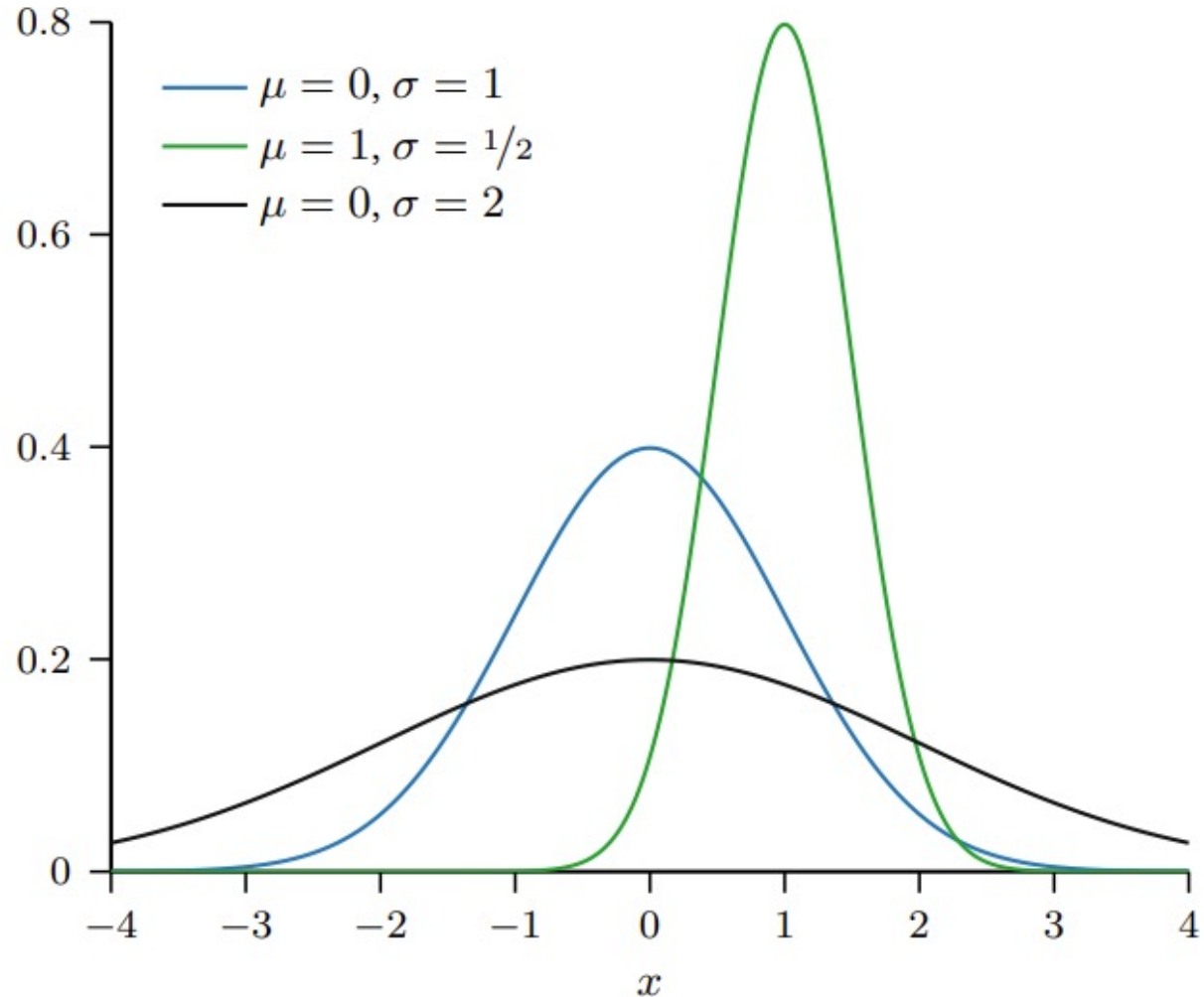
Then we can calculate the discrete probabilities for each class as before, once we have enough samples.

The Normal Distribution

Typically, though, we assume our continuous variables follow a Gaussian distribution.

The parameters μ and σ (mu and sigma) represent the mean and standard deviation respectively.

σ^2 is known as the variance.



The Normal Distribution

The probability density function of the univariate (one-dimensional) Gaussian distribution is

$$p(x \mid \mu, \sigma^2) = \mathcal{N}(x; \mu, \sigma^2) = \frac{1}{Z} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right).$$

The normalization constant Z is

$$Z = \sqrt{2\pi\sigma^2}.$$

The parameters μ and σ^2 specify the mean and variance of the distribution, respectively:

$$\mu = \mathbb{E}[x]; \quad \sigma^2 = \text{var}[x].$$

The Normal Distribution

Q: What has this go to do with Naïve Bayes again?

A: We use the normal distribution to calculate the probability of a class. To do this we will need to find the average and variance of the class samples.

$P(X=V|\text{class } A)$ reads “the probability of $X=V$ belonging to class A .” We can see this as a function. If we input temperature, x , into the function it will calculate and return the probability of that temperature belonging to class A .

$$f(x|\text{class } A) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

and the sample mean is:

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i$$

The standard deviation is σ :

$$\sigma = \sqrt{\frac{\sum (x - \mu)^2}{n - 1}}$$

The sample variance is:

$$\sigma^2 = \frac{\sum (x - \mu)^2}{n - 1}$$

Do I play Tennis today? Another example.

Given that it's **sunny**, the **temp is 66 degrees**, **humidity is 90** and it's **windy**, then should we play? Here's the table of data collected by our erstwhile player, let's call her Martina:-

outlook	temperature	humidity	windy	play
sunny	85	85	FALSE	no
sunny	80	90	TRUE	no
overcast	83	86	FALSE	yes
rainy	70	96	FALSE	yes
rainy	68	80	FALSE	yes
rainy	65	70	TRUE	no
overcast	64	65	TRUE	yes
sunny	72	95	FALSE	no
sunny	69	70	FALSE	yes
rainy	75	80	FALSE	yes
sunny	75	70	TRUE	yes
overcast	72	90	TRUE	yes
overcast	81	75	FALSE	yes
rainy	71	91	TRUE	no

How do we calculate the probability that we will play (or not)?

Step 1: Compute Probability of belonging to the temperature = 66 degrees for class Yes.

Step 1a: Calc mean of the temperature distribution for class Yes.

Step1b: Calc the variance of the temperature distribution for class Yes.

Step 1c: Calc the probability of temperature being 66 degrees for class Yes using the normal distribution formula.

Step 2: Compute Probability of belonging to the temperature = 66 degrees for class No (same method as step 1).

Do I play Tennis today?

Here's the table of data collected by our erstwhile player- let's call her Martina.

outlook	temperature	humidity	windy	play
sunny	85	85	FALSE	no
sunny	80	90	TRUE	no
overcast	83	86	FALSE	yes
rainy	70	96	FALSE	yes
rainy	68	80	FALSE	yes
rainy	65	70	TRUE	no
overcast	64	65	TRUE	yes
sunny	72	95	FALSE	no
sunny	69	70	FALSE	yes
rainy	75	80	FALSE	yes
sunny	75	70	TRUE	yes
overcast	72	90	TRUE	yes
overcast	81	75	FALSE	yes
rainy	71	91	TRUE	no

outlook	temperature	humidity	windy
sunny	66	90	TRUE

Step 1a: Calc mean of the temperature distribution for class Yes.

$$\mu \text{ (mean)} = (83+70+68+64+69+75+75+72+81)/9 = 73$$

Step1b: Calc the variance of the temperature distribution for class Yes.

$$\sigma^2 \text{ (variance)} = ((83-73)^2 + (70-73)^2 + (68-73)^2 + (64-73)^2 + (69-73)^2 + (75-73)^2 + (75-73)^2 + (72-73)^2 + (81-73)^2) / (9-1) = 38$$

Here's the table of data collected by our erstwhile player- let's call her Martina.

outlook	temperature	humidity	windy	play
sunny	85	85	FALSE	no
sunny	80	90	TRUE	no
overcast	83	86	FALSE	yes
rainy	70	96	FALSE	yes
rainy	68	80	FALSE	yes
rainy	65	70	TRUE	no
overcast	64	65	TRUE	yes
sunny	72	95	FALSE	no
sunny	69	70	FALSE	yes
rainy	75	80	FALSE	yes
sunny	75	70	TRUE	yes
overcast	72	90	TRUE	yes
overcast	81	75	FALSE	yes
rainy	71	91	TRUE	no

outlook	temperature	humidity	windy
sunny	66	90	TRUE

Step 1c: Calc the probability of temperature being 66 for class Yes using the normal distribution formula.

$$f(x|yes) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Where x= 66 degrees; μ = 73 degrees and σ = sqrt(38) degrees. Substitute into the formula:

$$f(x|yes) = \frac{1}{\sqrt{38*2*3.14}} 2.7^{-\frac{(x-73)^2}{2*38}}$$

$$f(x=66|yes) = \frac{1}{15.44} 2.7^{-\frac{(66-73)^2}{76}} = 0.034$$

$$P(\text{temp}=66|yes)=0.034$$

Here's the table of data collected by our erstwhile player- let's call her Martina.

outlook	temperature	humidity	windy	play
sunny	85	85	FALSE	no
sunny	80	90	TRUE	no
overcast	83	86	FALSE	yes
rainy	70	96	FALSE	yes
rainy	68	80	FALSE	yes
rainy	65	70	TRUE	no
overcast	64	65	TRUE	yes
sunny	72	95	FALSE	no
sunny	69	70	FALSE	yes
rainy	75	80	FALSE	yes
sunny	75	70	TRUE	yes
overcast	72	90	TRUE	yes
overcast	81	75	FALSE	yes
rainy	71	91	TRUE	no

outlook	temperature	humidity	windy
sunny	66	90	TRUE

Step 2: Compute Probability of the temperature = 66 degrees for class No.

Step 2a: Calc mean of the temperature distribution for class No.

Step2b: Calc the variance of the temperature distribution for class No.

Step 2c: Calc the probability of temperature being 66 for class No using the normal distribution formula.

Here's the table of data collected by our erstwhile player- let's call her Martina.

outlook	temperature	humidity	windy	play
sunny	85	85	FALSE	no
sunny	80	90	TRUE	no
overcast	83	86	FALSE	yes
rainy	70	96	FALSE	yes
rainy	68	80	FALSE	yes
rainy	65	70	TRUE	no
overcast	64	65	TRUE	yes
sunny	72	95	FALSE	no
sunny	69	70	FALSE	yes
rainy	75	80	FALSE	yes
sunny	75	70	TRUE	yes
overcast	72	90	TRUE	yes
overcast	81	75	FALSE	yes
rainy	71	91	TRUE	no

outlook	temperature	humidity	windy
sunny	66	90	TRUE

Step 2a: Calc mean of the temperature distribution for class No.

$$\mu \text{ (mean)} = (85+80+65+72+71)/5=74.6$$

Step2b: Calc the variance of the temperature distribution for class No.

$$\sigma^2 \text{ (variance)} = (85-75)^2 + (80-75)^2 + (65-75)^2 + (72-75)^2 + (71-75)^2 / (5-1) = 62.5$$

Here's the table of data collected by our erstwhile player- let's call her Martina.

outlook	temperature	humidity	windy	play
sunny	85	85	FALSE	no
sunny	80	90	TRUE	no
overcast	83	86	FALSE	yes
rainy	70	96	FALSE	yes
rainy	68	80	FALSE	yes
rainy	65	70	TRUE	no
overcast	64	65	TRUE	yes
sunny	72	95	FALSE	no
sunny	69	70	FALSE	yes
rainy	75	80	FALSE	yes
sunny	75	70	TRUE	yes
overcast	72	90	TRUE	yes
overcast	81	75	FALSE	yes
rainy	71	91	TRUE	no

outlook	temperature	humidity	windy
sunny	66	90	TRUE

Step 2c: Calc the probability of temperature being 66 for class No using the normal distribution formula.

$$f(x|no) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Where x= 66 degrees; μ = 74.6 degrees and σ = 7.9 degrees. Substitute into the formula:

$$P(\text{temp}=66 | \text{no}) = \frac{1}{\sqrt{2\pi}7.9} e^{-\frac{(66-74.6)^2}{2*62.3}} = 0.02791$$

Step3: Calculate the prior probabilities of a yes and a no.

$$P(\text{play=yes})=9/14; P(\text{play=no})=5/14$$

Do I play Tennis today?

Now repeat for the humidity class for a yes distribution and a no distribution.

- $P(\text{humidity}=90|\text{yes})=0.022$; $P(\text{humidity}=90|\text{no})=0.038$

Step 4: Next we calculate the Yes and No probabilities for the remaining discrete classes (as we did with Lara- remember?).

- $P(\text{outlook}=\text{sunny}|\text{yes})=2/9$; $P(\text{outlook}=\text{sunny}|\text{no})=3/5$
- $P(\text{windy}=\text{yes}|\text{yes})=3/9$; $P(\text{windy}=\text{yes}|\text{no})=3/5$

Step 5: Finally we calculate the probability of playing given the above conditional probabilities for a Yes then repeat for a No. What is the calc for Yes, I will play? It's:-

$$P(\text{yes}) * P(\text{outlook}=\text{sunny}|\text{yes}) * P(\text{temp}=66|\text{yes}) * P(\text{humidity}=90|\text{yes}) * P(\text{windy}=\text{yes}|\text{yes})$$

$$9/14 * 2/9 * 0.034 * 0.022 * 3/9 = 0.000036$$

And for No, I won't play tennis? It's:-

$$P(\text{no}) * P(\text{outlook}=\text{sunny}|\text{no}) * P(\text{temp}=66|\text{no}) * P(\text{humidity}=90|\text{no}) * P(\text{windy}=\text{yes}|\text{no})$$

$$5/14 * 3/5 * 0.027918 * 0.038 * 0.6 = 0.000136$$

So No I won't play as $P(\text{no Play}|\dots) > P(\text{Play}|\dots)$ Easy!

Issues with Naïve Bayes

- Violation of Independence Assumption
 - For many real world tasks, $P(X_1, \dots, X_n | C) \neq P(X_1 | C) \dots P(X_n | C)$
 - Nevertheless, naïve Bayes works surprisingly well anyway!
- Zero conditional probability Problem
 - If no example contains the attribute value $X_j = a_{jk}$, $\hat{P}(X_j = a_{jk} | C = c_i) = 0$
 - In this circumstance, $\hat{P}(x_1 | c_i) \dots \hat{P}(a_{jk} | c_i) \dots \hat{P}(x_n | c_i) = 0$ during test

Advantages of Naïve Bayes

- Very simple and easy to implement.
- It needs less training data.
- It is highly scalable with number of predictors and data points.
- It is fast classification algorithm.
- It is not sensitive to irrelevant features.