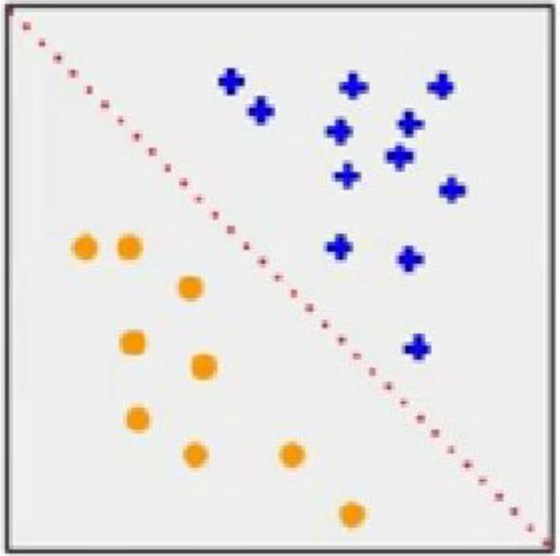
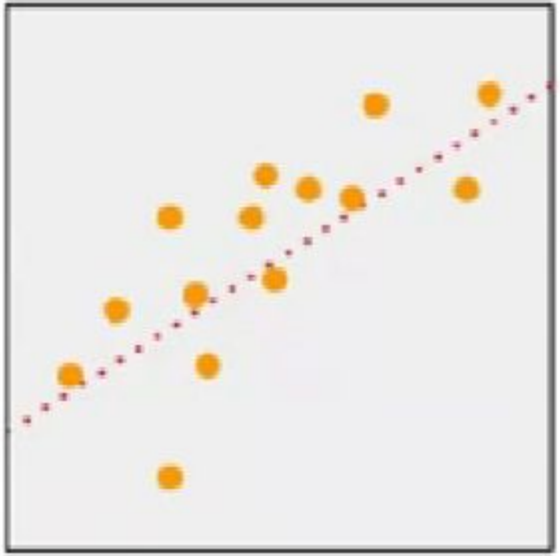
Naive bayes classification

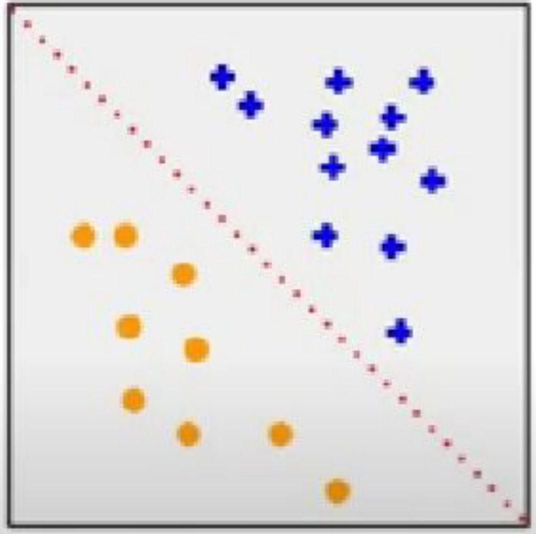
Bayes Classifier



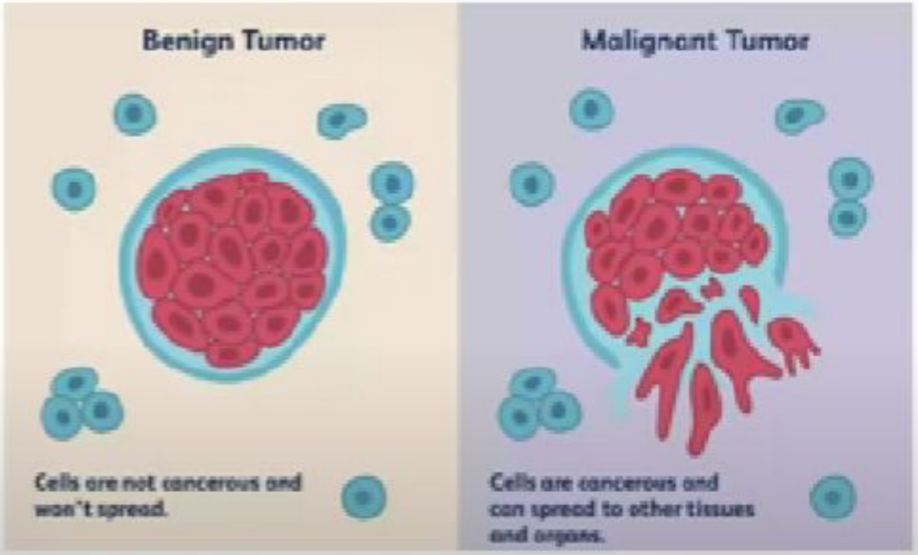
Classification

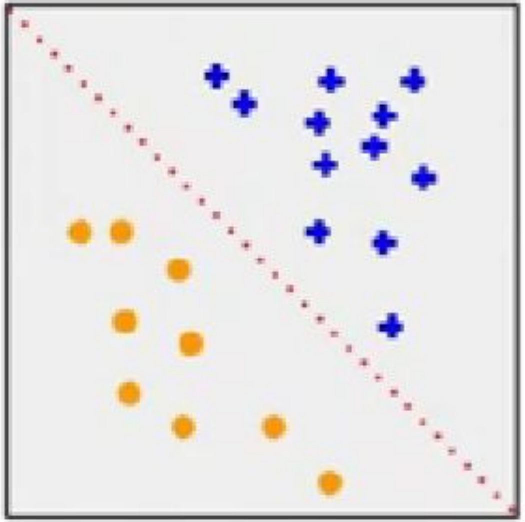


Regression

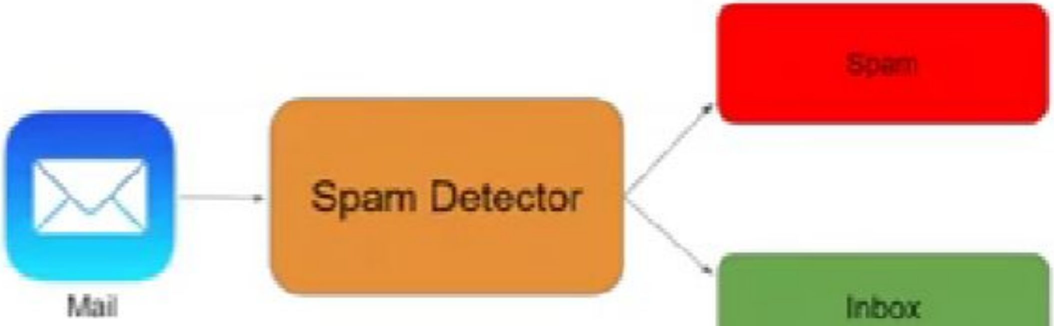


Classification

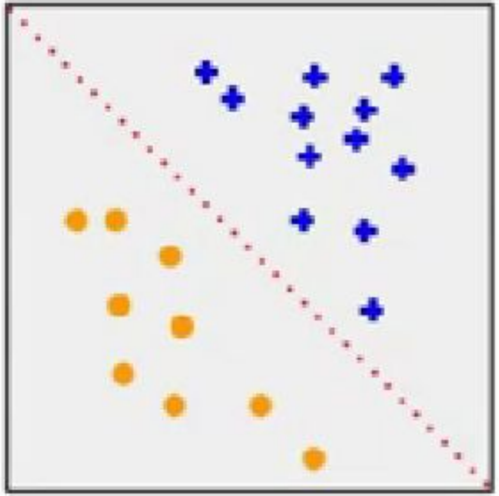




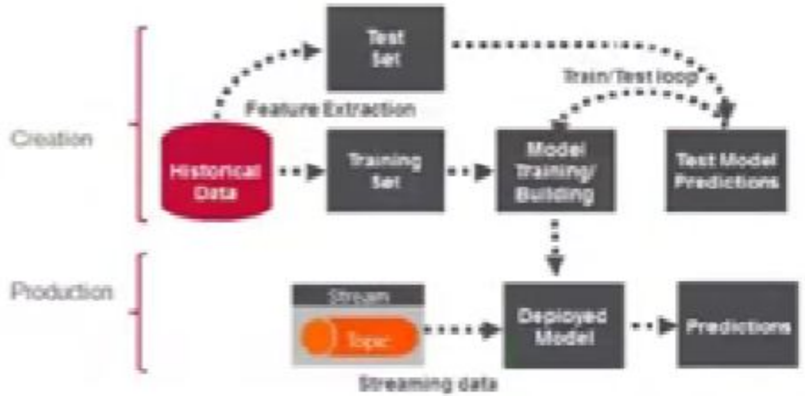
Classification



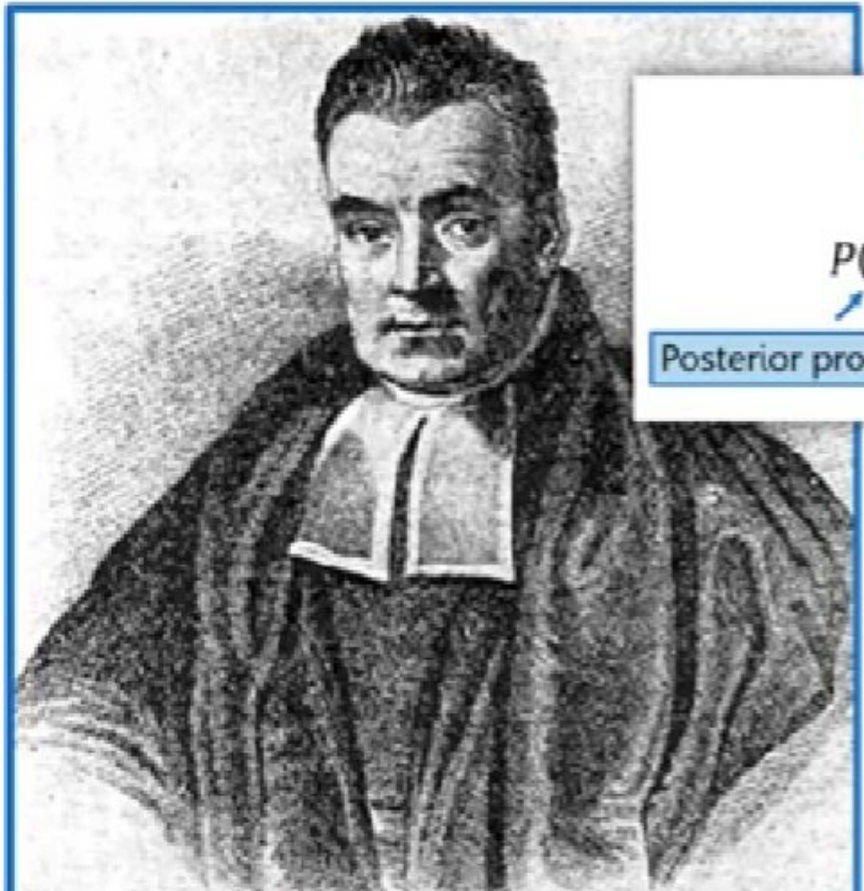
Credit Card Fraud Detection



Classification



Bayes Classifier



Likelihood

Prior probability

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

Posterior probability

Evidence

Bayes Classifier

- Bayes Rule-

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

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

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

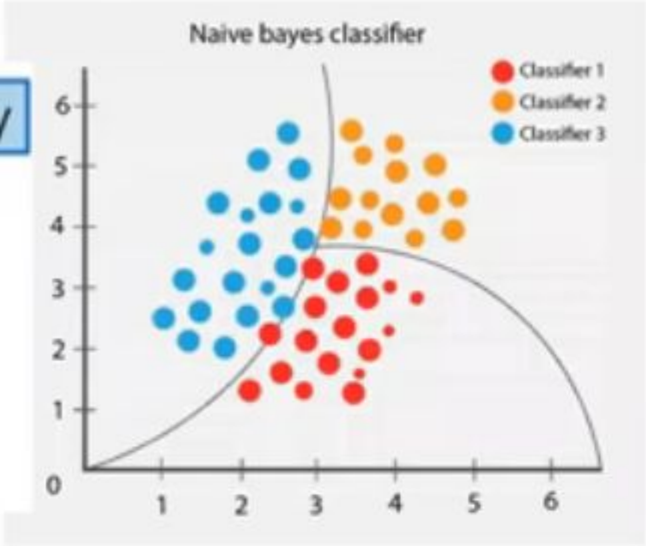
$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Bayes Classifier

- Bayes Rule- $P(A|B) = \dfrac{P(B|A)\,P(A)}{P(B)}$

In machine learning, naive Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naive) independence assumptions between the features.

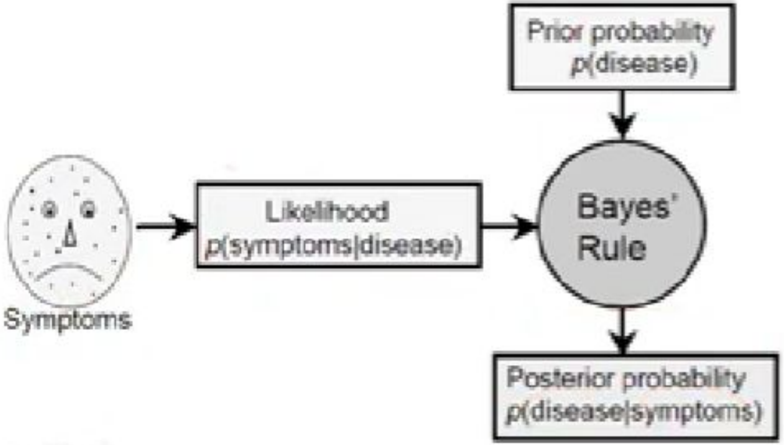
$$\underbrace{P(A|B)}_{\text{Posterior probability}} = \frac{\overbrace{P(B|A)}^{\text{Likelihood}} \cdot \overbrace{P(A)}^{\text{Prior probability}}}{\underbrace{P(B)}_{\text{Evidence}}}$$



Bayes Classifier- p(smallpox|spots)?

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

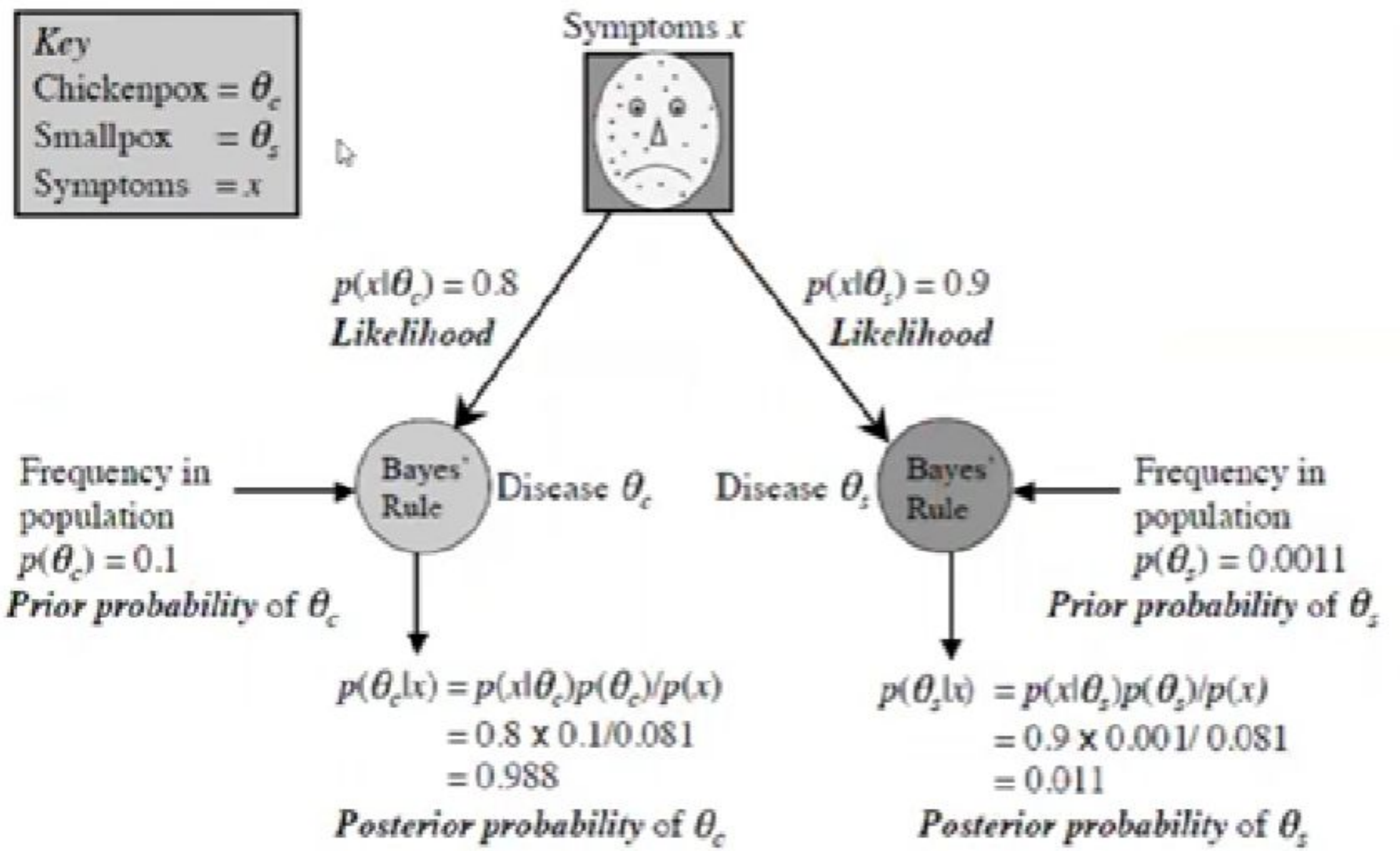
$$p(\text{smallpox}|\text{spots}) = \frac{p(\text{spots}|\text{smallpox}) \times p(\text{smallpox})}{p(\text{spots})}$$

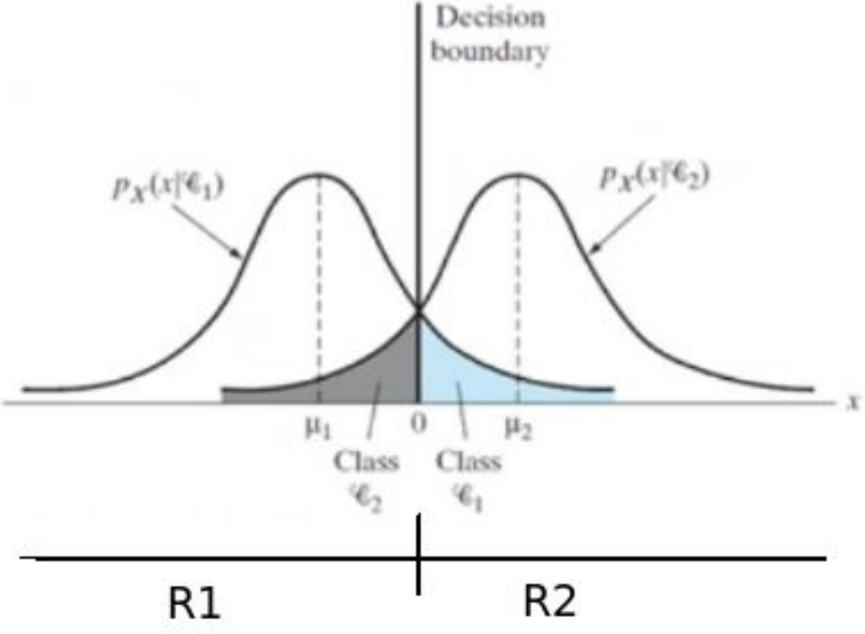


Example-

$$p(\text{smallpox}) = 0.001$$
$$p(\text{spots}|\text{smallpox}) = 0.9$$
$$p(\text{spots}) = 0.081$$
$$\begin{aligned} p(\text{smallpox}|\text{spots}) &= 0.9 \times 0.001/0.081 \\ &= 0.011, \end{aligned}$$

Bayes Classifier- p(smallpox|spots)?





How many misclassifications happened for Class H1?

Errors are committed due to region R2 colored in light blue and is given by

$$p(H1) \int_{R2} p(x/H1)\, dx$$

Bayes Classifier- To minimize the average risk

$$\mathcal{R} = c_{11}p_1\int_{\mathcal{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1)d\mathbf{x} + c_{22}p_2\int_{\mathcal{X}_2} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2)d\mathbf{x} + c_{21}p_1\int_{\mathcal{X}_2} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1)d\mathbf{x} + c_{12}p_2\int_{\mathcal{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2)d\mathbf{x}$$

minimize Average Risk

Design a classifier

Email 100 → Gmail: Spam classifier

Class 1: Spam Folder 70

Class 2: Inbox 30

Assigned class \ Actual class		
	C_{11} Correct Decision	C_{12} Incorrect decision
	C_{21} Incorrect	C_{22} Correct

C_{11}: Assign Spam, Actual Spam

C_{12}: Assign Spam, Actual Inbox

C_{ij}

21: Assigned class Inbox, Actual Class Spam

Bayes Classifier- To minimize the average risk

$$\mathcal{R} = c_{11}p_1 \int_{\mathcal{H}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1)d\mathbf{x} + c_{22}p_2 \int_{\mathcal{H}_2} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2)d\mathbf{x} + c_{21}p_1 \int_{\mathcal{H}_2} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1)d\mathbf{x} + c_{12}p_2 \int_{\mathcal{H}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2)d\mathbf{x}$$

Each decision is weighted by the product of two factors- the cost involved in making the decision and the relative frequency (i.e., prior probability) with which it occurs

First two terms on the right-hand side represent correct decisions/classifications, whereas the last two terms represent incorrect decisions

p_i = *prior probability* that the observation vector $\mathbf{x}$ (representing a realization of the random vector $\mathbf{X}$) corresponds to an object in class C_1, with $i = 1, 2$, and $p_1 + p_2 = 1$

c_{ij} = cost of deciding in favor of class $\mathcal{C}_i$ represented by subspace $\mathcal{H}_i$ when class $\mathcal{C}_j$ is true (i.e., observation vector $\mathbf{x}$ corresponds to an object in class C_1), with $i, j = 1, 2$

$p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_i)$ = conditional probability density function of the random vector $\mathbf{X}$, given that the observation vector $\mathbf{x}$ corresponds to an object in class C_1, with $i = 1, 2$.

Bayes Classifier- To minimize the average risk

$$\mathcal{R} = c_{11}p_1\int_{\mathcal{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1)d\mathbf{x} + c_{22}p_2\int_{\mathcal{X}_2} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2)d\mathbf{x} + c_{21}p_1\int_{\mathcal{X}_2} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1)d\mathbf{x} + c_{12}p_2\int_{\mathcal{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2)d\mathbf{x}$$

The intention is to determine a strategy for the *minimum average risk*. Because we require that a decision be made, each observation vector **x** must be assigned in the overall observation space $\mathcal{X}$ to either $\mathcal{X}_1$ or $\mathcal{X}_2$. Thus,

Bayes Classifier- To minimize the average risk

$$\mathcal{R} = c_{11}p_1 \int_{\mathcal{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1)d\mathbf{x} + c_{22}p_2 \int_{\mathcal{X}_2} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2)d\mathbf{x} + c_{21}p_1 \int_{\mathcal{X}_2} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1)d\mathbf{x} + c_{12}p_2 \int_{\mathcal{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2)d\mathbf{x}$$

$$\mathcal{R} = c_{11}p_1 \int_{\mathcal{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1)d\mathbf{x} + c_{22}p_2 \int_{\mathcal{X}-\mathcal{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2)d\mathbf{x} + c_{21}p_1 \int_{\mathcal{X}-\mathcal{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1)\mathbf{x} + c_{12}p_2 \int_{\mathcal{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2)d\mathbf{x}$$

We now observe the fact that-

$$\int_{\mathcal{X}} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1)d\mathbf{x} = \int_{\mathcal{X}} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2)d\mathbf{x} = 1$$

Bayes Classifier- To minimize the average risk

$$\mathcal{R} = c_{11}p_1 \int_{\mathcal{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1)d\mathbf{x} + c_{22}p_2 \int_{\mathcal{X}_2} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2)d\mathbf{x} + c_{21}p_1 \int_{\mathcal{X}_2} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1)d\mathbf{x} + c_{12}p_2 \int_{\mathcal{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2)d\mathbf{x}$$

$$\mathcal{R} = c_{11}p_1 \int_{\mathcal{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1)d\mathbf{x} + c_{22}p_2 \int_{\mathcal{X}-\mathcal{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2)d\mathbf{x} + c_{21}p_1 \int_{\mathcal{X}-\mathcal{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1)\mathbf{x} + c_{12}p_2 \int_{\mathcal{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2)d\mathbf{x}$$

$$\mathcal{R} = c_{21}p_1 + c_{22}p_2 + \int_{\mathcal{X}_1} [p_2(c_{12} - c_{22})\, p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2) - p_1(c_{21} - c_{11})\, p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1)]d\mathbf{x}$$

Bayes Classifier- To minimize the average risk

$$\mathscr{R} = c_{11}p_1 \int_{\mathscr{H}_1} p_{\mathbf{X}}(\mathbf{x}|\mathscr{C}_1)d\mathbf{x} + c_{22}p_2 \int_{\mathscr{H}_2} p_{\mathbf{X}}(\mathbf{x}|\mathscr{C}_2)d\mathbf{x} + c_{21}p_1 \int_{\mathscr{H}_2} p_{\mathbf{X}}(\mathbf{x}|\mathscr{C}_1)d\mathbf{x} + c_{12}p_2 \int_{\mathscr{H}_1} p_{\mathbf{X}}(\mathbf{x}|\mathscr{C}_2)d\mathbf{x}$$

$$\mathscr{R} = c_{11}p_1 \int_{\mathscr{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathscr{C}_1)d\mathbf{x} + c_{22}p_2 \int_{\mathscr{X}-\mathscr{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathscr{C}_2)d\mathbf{x} + c_{21}p_1 \int_{\mathscr{X}-\mathscr{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathscr{C}_1)\mathbf{x} + c_{12}p_2 \int_{\mathscr{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathscr{C}_2)d\mathbf{x}$$

$$\mathscr{R} = c_{21}p_1 + c_{22}p_2 + \int_{\mathscr{X}_1} \boxed{[p_2(c_{12} - c_{22})\, p_{\mathbf{X}}(\mathbf{x}|\mathscr{C}_2) - p_1(c_{21} - c_{11})\, p_{\mathbf{X}}(\mathbf{x}|\mathscr{C}_1)]} d\mathbf{x}$$

$$p_1(c_{21} - c_{11})\, p_{\mathbf{X}}(\mathbf{x}|\mathscr{C}_1) > p_2(c_{12} - c_{22})\, p_{\mathbf{X}}(\mathbf{x}|\mathscr{C}_2)$$

Bayes Classifier- To minimize the average risk

$$\mathcal{R} = c_{11}p_1 \int_{\mathcal{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1)d\mathbf{x} + c_{22}p_2 \int_{\mathcal{X}_2} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2)d\mathbf{x} + c_{21}p_1 \int_{\mathcal{X}_2} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1)d\mathbf{x} + c_{12}p_2 \int_{\mathcal{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2)d\mathbf{x}$$

$$\mathcal{R} = c_{11}p_1 \int_{\mathcal{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1)d\mathbf{x} + c_{22}p_2 \int_{\mathcal{X}-\mathcal{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2)d\mathbf{x} + c_{21}p_1 \int_{\mathcal{X}-\mathcal{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1)\mathbf{x} + c_{12}p_2 \int_{\mathcal{X}_1} p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2)d\mathbf{x}$$

$$\mathcal{R} = c_{21}p_1 + c_{22}p_2 + \int_{\mathcal{X}_1} \boxed{[p_2(c_{12} - c_{22})\, p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2) - p_1(c_{21} - c_{11})\, p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1)]} d\mathbf{x}$$

$$p_1(c_{21} - c_{11})\, p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1) > p_2(c_{12} - c_{22})\, p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2)$$

Bayes Classifier Rule-

If, for an observation vector $\mathbf{x}$*, the likelihood ratio* $\Lambda(\mathbf{x})$ *is greater than the threshold* ξ*, assign* $\mathbf{x}$ *to class* $\mathcal{C}_1$*. Otherwise, assign it to class* $\mathcal{C}_2$*.*

Let us define-

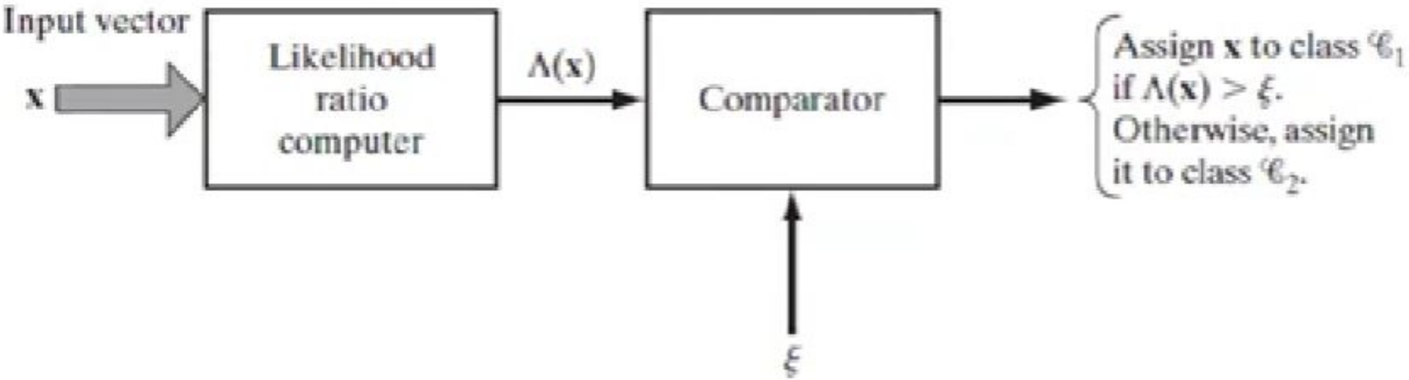
Likelihood Ratio

$$\Lambda(\mathbf{x}) = \frac{p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_1)}{p_{\mathbf{X}}(\mathbf{x}|\mathcal{C}_2)}$$

Threshold

$$\xi = \frac{p_2(c_{12} - c_{22})}{p_1(c_{21} - c_{11})}$$

Bayes Classifier- To minimize the average risk



$$p_1(c_{21} - c_{11})\, p_X(\mathbf{x}|\mathcal{C}_1) > p_2(c_{12} - c_{22})\, p_X(\mathbf{x}|\mathcal{C}_2)$$

Bayes Classifier Rule-

If, for an observation vector $\mathbf{x}$, *the likelihood ratio* $\Lambda(\mathbf{x})$ *is greater than the threshold* ξ, *assign* $\mathbf{x}$ *to class* $\mathcal{C}_1$. *Otherwise, assign it to class* $\mathcal{C}_2$.

Let us define-

Likelihood Ratio

$$\Lambda(\mathbf{x}) = \frac{p_X(\mathbf{x}|\mathcal{C}_1)}{p_X(\mathbf{x}|\mathcal{C}_2)}$$

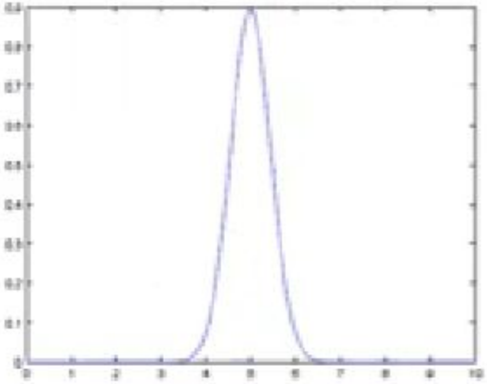
Threshold

$$\xi = \frac{p_2(c_{12} - c_{22})}{p_1(c_{21} - c_{11})}$$

Multivariate Gaussian Distribution

Univariate Gaussian Function-

$$p(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$



Gauss

(30 April 1777 – 23 February 1855)
German mathematician and physicist

Multivariate Gaussian Distribution

$$p(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

Consider m=2-variable case-

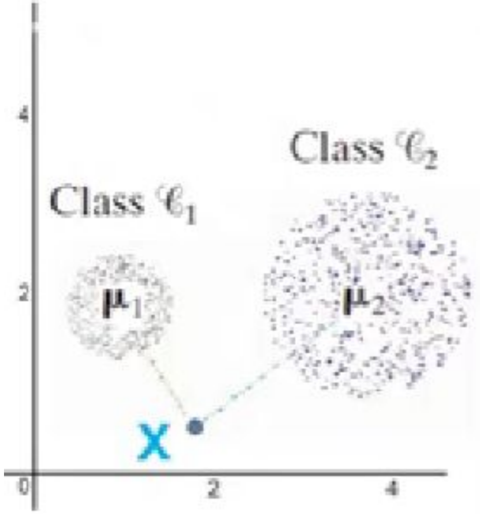
$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

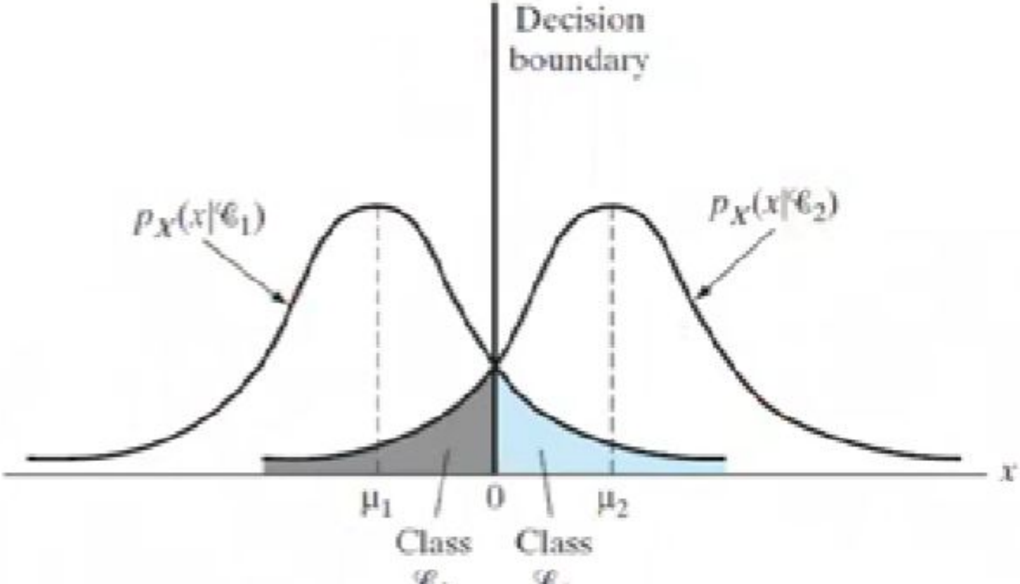
$$p_{\mathbf{X}}(\mathbf{x}|\mathscr{C}_i) = \frac{1}{(2\pi)^{m/2}(\det(\mathbf{C}))^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T\mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right)$$

$$\begin{aligned}
p(x;\mu,\Sigma) &= \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{1}{2\sigma_1^2}(x_1-\mu_1)^2\right)\cdot\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left(-\frac{1}{2\sigma_2^2}(x_2-\mu_2)^2\right) \\
&= \frac{1}{2\pi\sigma_1\sigma_2}\exp\left(-\frac{1}{2\sigma_1^2}(x_1-\mu_1)^2-\frac{1}{2\sigma_2^2}(x_2-\mu_2)^2\right) \\
&= \frac{1}{2\pi\sigma_1\sigma_2}\exp\left(-\frac{1}{2}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}^T\begin{bmatrix} \frac{1}{\sigma_1^2}(x_1-\mu_1) \\ \frac{1}{\sigma_2^2}(x_2-\mu_2) \end{bmatrix}\right) \\
&= \frac{1}{2\pi\begin{vmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{vmatrix}^{1/2}}\exp\left(-\frac{1}{2}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}^T\begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}^{-1}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}\right)
\end{aligned}$$

Bayes Classifier in Gaussian Environment

$$\boxed{p_{\mathbf{X}}(\mathbf{x}|\mathscr{C}_i) = \frac{1}{(2\pi)^{m/2}(\det(\mathbf{C}))^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right)}$$





Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

Frequency and Likelihood tables of ‘Color’

Frequency Table

		Stolen?	
		Yes	No
Color	Red	3	2
	Yellow	2	3

Likelihood Table

		Stolen?	
		P(Yes)	P(No)
Color	Red	3/5	2/5
	Yellow	2/5	3/5

Frequency and Likelihood tables of ‘Type’

Frequency Table

		Stolen?	
		Yes	No
Type	Sports	4	2
	SUV	1	3

Likelihood Table

		Stolen?	
		P(Yes)	P(No)
Type	Sports	4/5	2/5
	SUV	1/5	3/5

Frequency and Likelihood tables of ‘Origin’

Frequency Table

		Stolen?	
		Yes	No
Origin	Domestic	2	3
	Imported	3	2

Likelihood Table

		Stolen?	
		P(Yes)	P(No)
Origin	Domestic	2/5	3/5
	Imported	3/5	2/5

Color	Type	Origin	Stolen
Red	SUV	Domestic	**?**

$$P(Yes \mid X) = P(Red \mid Yes) * P(SUV \mid Yes) * P(Domestic \mid Yes) * P(Yes)$$

$$= \frac{3}{5} * \frac{1}{5} * \frac{2}{5} * 0.5$$

$$= 0.048 * 0.5$$

$$P(No \mid X) = P(Red \mid No) * P(SUV \mid No) * P(Domestic \mid No) * P(No)$$

$$= \frac{2}{5} * \frac{3}{5} * \frac{3}{5} * 0.5$$

$$= 0.144 * 0.5$$

Since 0.144 > 0.048, Which means given the features RED SUV and Domestic, our example gets classified as ’NO’ the car is not stolen.

Example

Name	Yellow	Sweet	Long
Mango	350	450	0
Banana	400	300	350
Others	50	100	50

Step 1-: frequency table

Name	Yellow	Sweet	Long	Total
Mango	350	450	0	650
Banana	400	300	350	400
Others	50	100	50	150
Total	800	850	400	1200

Step 2: Draw the likelihood table for the features against the classes.

Name	Yellow	Sweet	Long	Total
Mango	350/800=P(Mango|Yellow)	450/850	0/400	650/1200=P(Mango)
Banana	400/800	300/850	350/400	400/1200
Others	50/800	100/850	50/400	150/1200
Total	800=P(Yellow)	850	400	1200

Classify the fruit which is long,sweet and yellow

Step 4:

- Calculate . In our example, the maximum probability is for the class banana, therefore, the fruit which is long, sweet and yellow is a banana by Naive Bayes Algorithm.

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

$$P(PlayTennis = yes) = 9/14 = .64$$

$$P(PlayTennis = no) = 5/14 = .36$$

Outlook	Y	N
sunny	2/9	3/5
overcast	4/9	0
rain	3/9	2/5
Tempreature		
hot	2/9	2/5
mild	4/9	2/5
cool	3/9	1/5

Humidity	Y	N
high	3/9	4/5
normal	6/9	1/5
Windy		
Strong	3/9	3/5
Weak	6/9	2/5

NAIVE BAYES CLASSIFIER
Example - 1