CS4801 : Bayesian Decision Theory

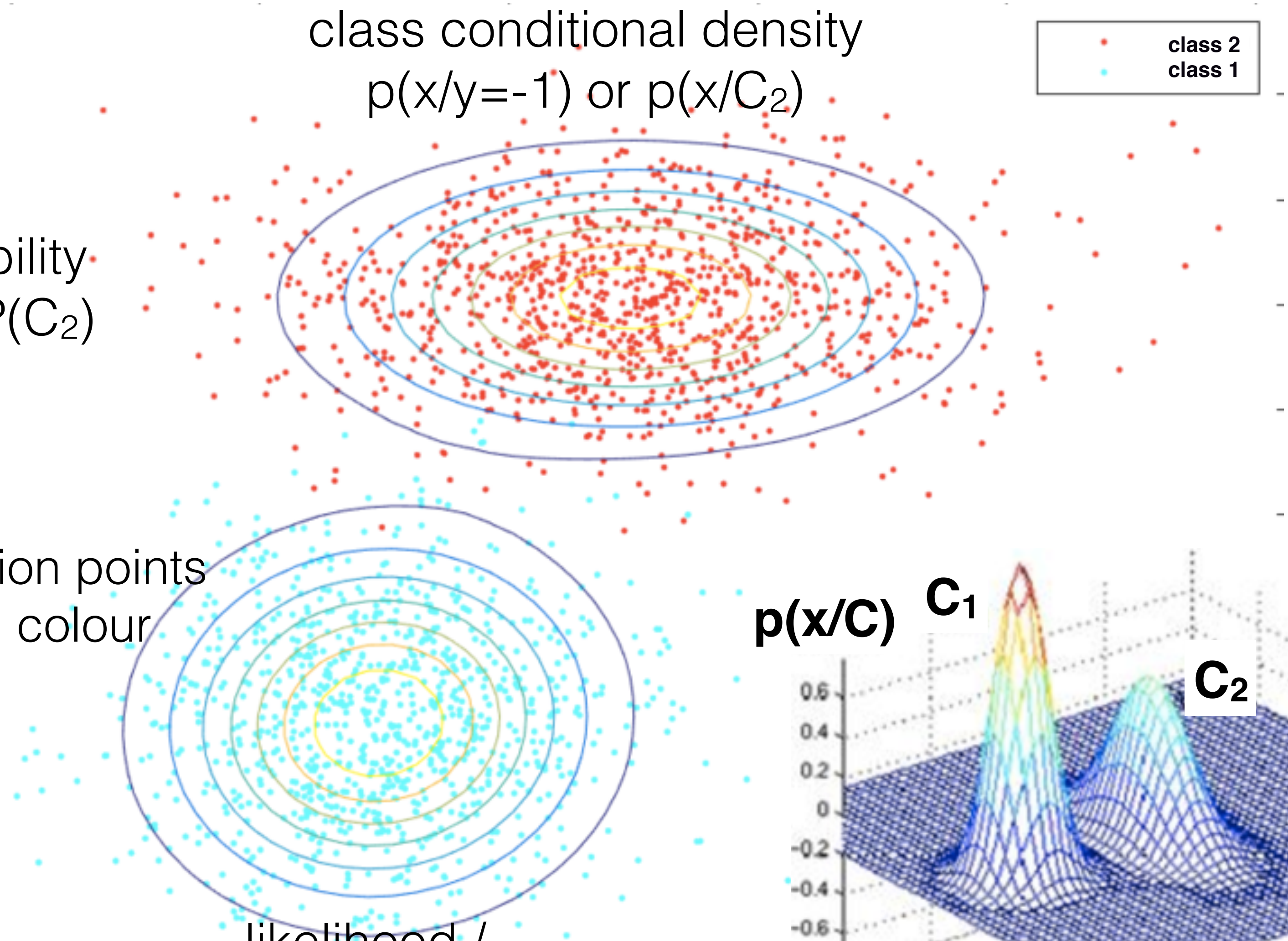
Sahely Bhadra
16/8/2017

2.Bayes decision rule
   1. Classification error
   2. Minimum error rate classification
      1. Two category
      2. Multi category

Probabilistic Decision Boundary

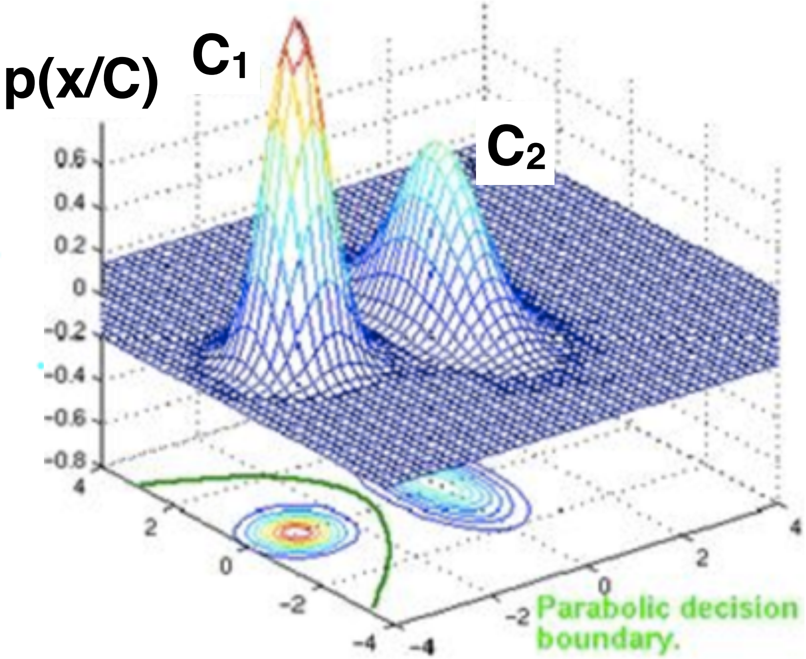
likelihood /
class conditional density
$p(x/y=-1)$ or $p(x/C_2)$

prior probability
$P(C_1)$ and $P(C_2)$



$P(C_1)$= fraction points having cyan colour

likelihood /
class conditional density
$p(x/y=1)$ or $p(x/C_1)$



Bayes Classifier

New point (x_t)

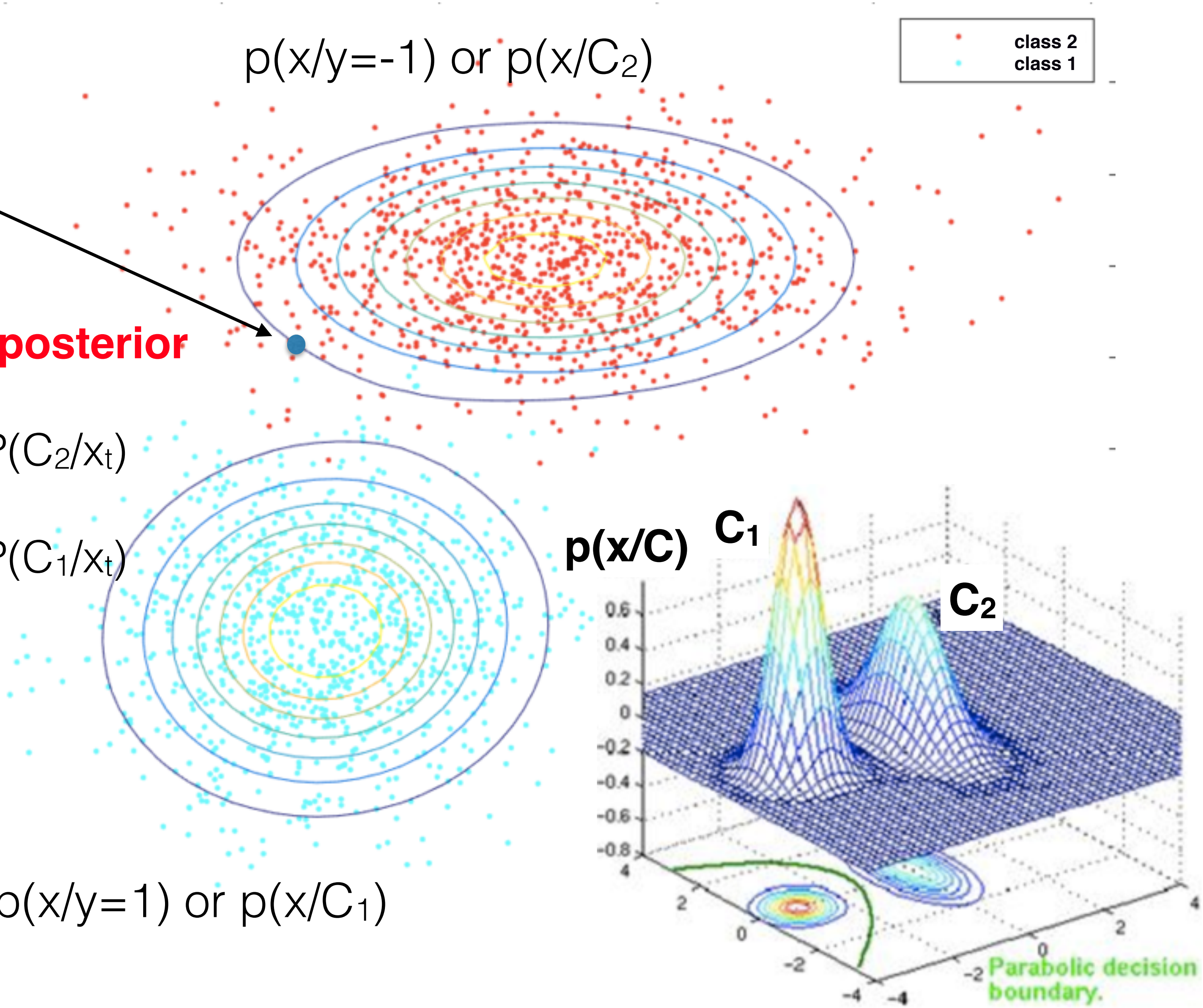
Classification on posterior

Label =1

if $P(C_1/x_t) > P(C_2/x_t)$

Label =-1

if $P(C_2/x_t) > P(C_1/x_t)$



Bayes Classifier

Classification on posterior

Label =1

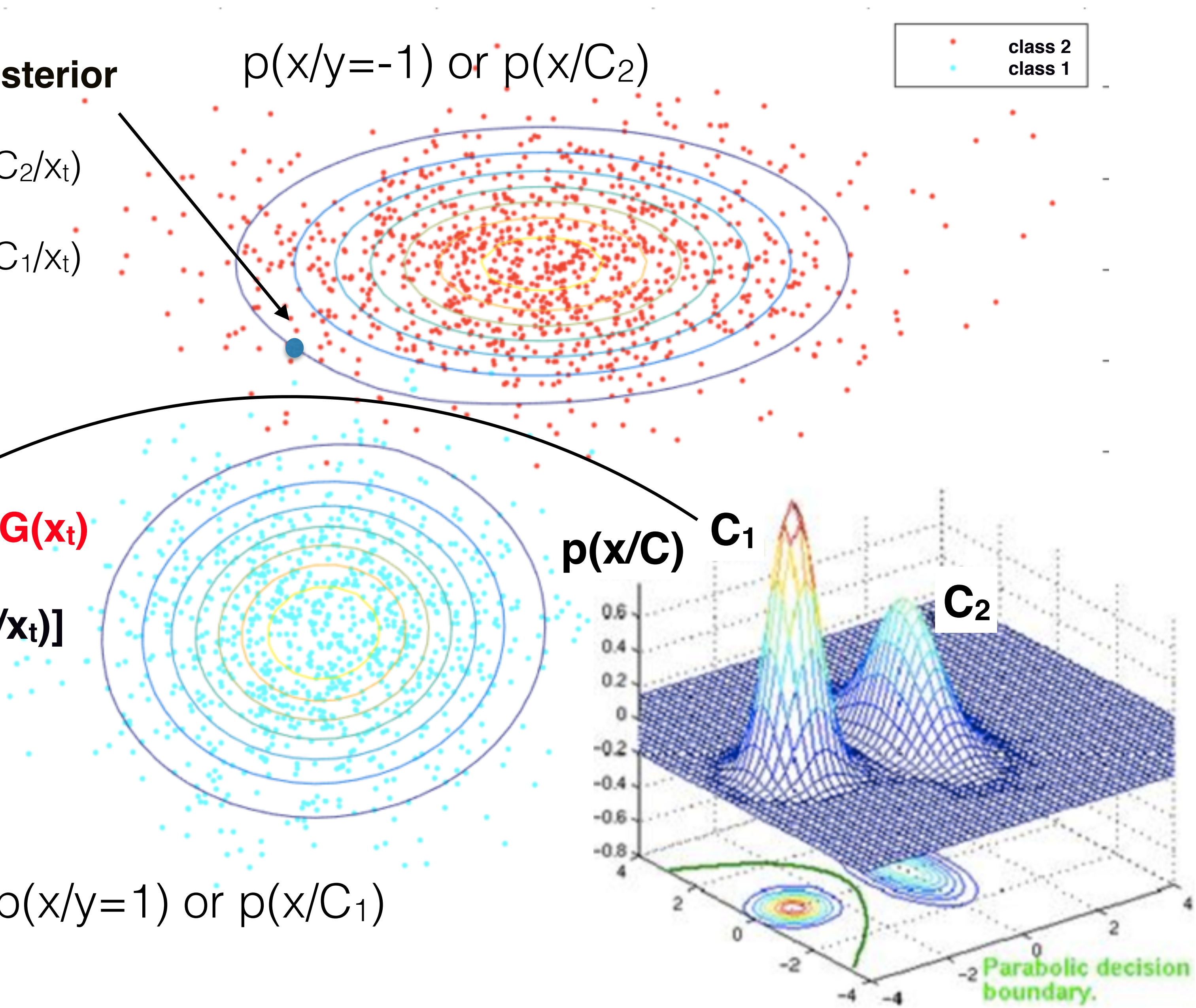
if $P(C_1/x_t) > P(C_2/x_t)$

Label =-1

if $P(C_2/x_t) > P(C_1/x_t)$

Decision Function G(x_t)

sign[$P(C_1/x_t) - P(C_2/x_t)$]



Probabilistic Decision Boundary

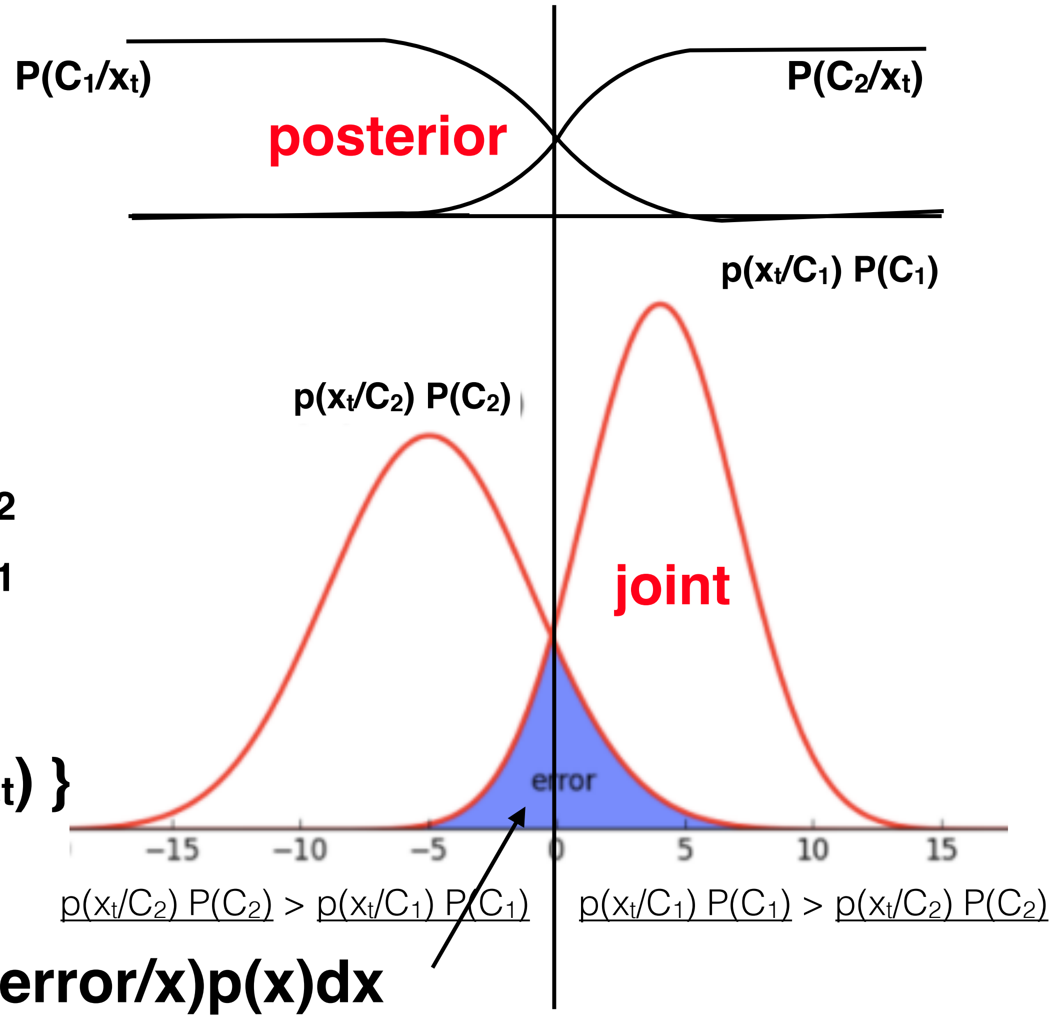
Classification on posterior

Label =1

if $P(C_1/x_t) > P(C_2/x_t)$

Label =-1

if $P(C_2/x_t) > P(C_1/x_t)$



$$P(error/x) = \begin{cases} P(C_1/x_t) & \text{if decide } C_2 \\ P(C_2/x_t) & \text{if decide } C_1 \end{cases}$$

$$P(error/x) = \min \{ P(C_1/x_t), P(C_2/x_t) \}$$

$$P(error) = \int_{-\infty}^{\infty} P(error, x)dx = \int_{-\infty}^{\infty} P(error/x)p(x)dx$$

The Bayes rule is optimum, that is, it minimises the average probability error!

More general theory

- Use more than one features. $[x_1, x_2, \ldots, x_d]$
- Allow more than two categories. $[C_1, C_2, \ldots C_k]$
- Employ a more general error function (i.e., "**risk**" function) by associating a "cost" ("loss" function) with each error (i.e., wrong action).

- $L_{ij} = cost(A_i / C_j)$
- = cost of incorrectly taking action "i" (A_i) when the correction action is "j" (C_j)

The conditional risk (or **expected loss**) with taking action (A_i)

$$\mathbf{R(A_i/x) = \sum_{j=1}^{k} cost(A_i / C_j) P(C_j /x)}$$

Overall Risk (R) $= \int_{-\infty}^{X} \mathbf{R(A(x)/x)\ p(X)\ dx}$

Bayes Decision Rule

$$\textbf{Overall Risk} \ (R) = \int_X R(A(x)/x)$$

The Bayes decision rule minimises R by:

(i) Computing $R(A_i/x)$ for every (A_i) given an x

(ii) Choosing the action (A_i) with the minimum $R(A_i/x)$

The resulting minimum overall risk is called **Bayes risk** and is the **best** (i.e., optimum) performance that can be achieved

Naive Bayes

- Use more than one features. $[x_1,x_2,...,x_d]$
- Features are independent

FLU	chills	headache	fever
N	Y	M	Y
Y	Y	N	N
Y	Y	S	Y
Y	N	M	Y
N	N	N	N
Y	N	S	Y
N	N	S	N
Y	Y	M	Y
?	Y	M	N

P(Ch =Y/Flu =Y)=3/5
P(Ch =N/Flu =Y)=2/5
P(Ch =Y/Flu =N)=1/3
P(Ch =N/Flu =N)=2/3

P(Ha=M/Flu=Y)=2/5
P(Ha=N/Flu=Y)=1/5
P(Ha=S/Flu=Y)=2/5
P(Ha=M/Flu=N)=1/3
P(Ha=N/Flu=N)=1/3
P(Ha=S/Flu=N)=1/3

P(Fe=Y/Flu=Y)=4/5
P(Fe=N/Flu=Y)=1/5
P(Fe=Y/Flu=N)=1/3
P(Fe=N/Flu=N)=2/3

P(Flu=Y)=5/8
P(Flu=N) =3/8

z=P(Ch =Y,Ha=M,Fn=N)

P(Flu=Y/Ch =Y,Ha=M,Fn=N)
= P(Ch =Y,Ha=M,Fn=N /Flu=Y) P(Flu=Y)/z
= P(Ch =Y/Flu=Y)P(Ha=M//Flu=Y)P(Fn=N /Flu=Y)P(Flu=Y)/z
=3/5*2/5*1/5 *5/8=3/100z

P(Flu=N/Ch =Y,Ha=M,Fn=N)
= P(Ch =Y,Ha=M,Fn=N /Flu=N) P(Flu=N)/z
= P(Ch =Y/Flu=N)P(Ha=M//Flu=N)P(Fn=N /Flu=N)P(Flu=N)/z
=1/3*1/3*2/3 *3/8=1/36z

P(Flu=Y/Ch =Y,Ha=M,Fn=N) > P(Flu=N/Ch =Y,Ha=M,Fn=N)

Next Class

- 21/8
  - Logistic Regression
  - Parzen window : density estimation
  - Perceptron classifier
- 22/8
  - short quite + solving assignment