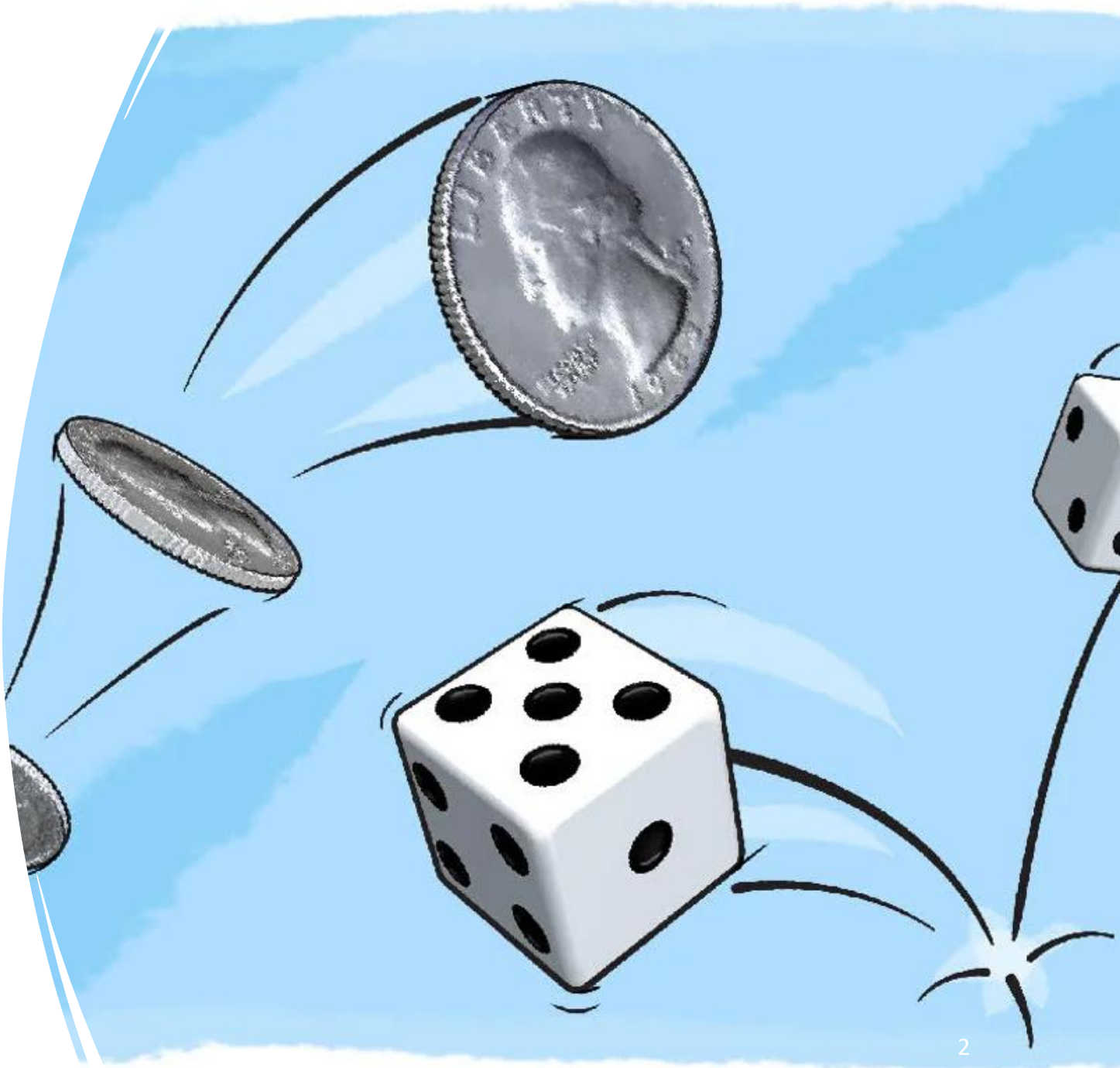
ML - 3

Machine Learning

- Supervised Learning
  - Probability based Classification
    - Naïve Bayes Classifier

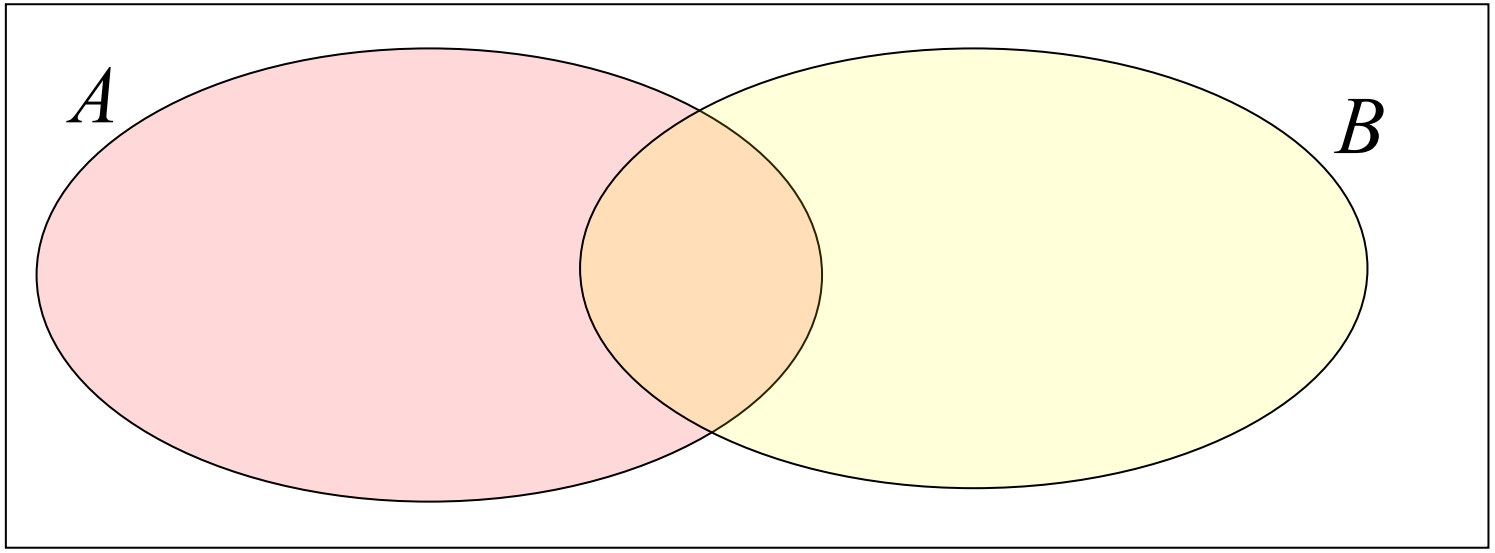
Probability

The probability of a particular event is the frequency of occurrence of that event out of all possible outcomes



Conditional Probability

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



Bayes Theorem

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

$$P(A \cap B) = P(A \mid B)P(B) = P(B \mid A)P(A)$$

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

Example-1

Suppose the probability of any person in a city having cancer is 0.8%, and we have a test to detect cancer. It is known that the probability of having cancer is 0.98 for individuals who test positive (+), while the probability of being healthy is 0.97 for those who test negative (−).

$$P(cancer) = 0.008 \qquad P(\neg cancer) = 0.992$$

$$P(+ \mid cancer) = 0.98 \qquad P(+ \mid \neg cancer) = 0.03$$

$$P(- \mid cancer) = 0.02 \qquad P(- \mid \neg cancer) = 0.97$$

Example-1

$$P(cancer \mid +) = \frac{0.98 \times 0.008}{P(+) = 0.0376} = 0.2085$$

$$P(\neg cancer \mid +) = \frac{0.03 \times 0.992}{P(+) = 0.0376} = 0.7915$$

$$P(cancer \mid +) < P(\neg cancer \mid +)$$

Maximum a posteriori

$$P(cancer \mid +) = \frac{0.98 \times 0.008}{P(+) = 0.0376}$$

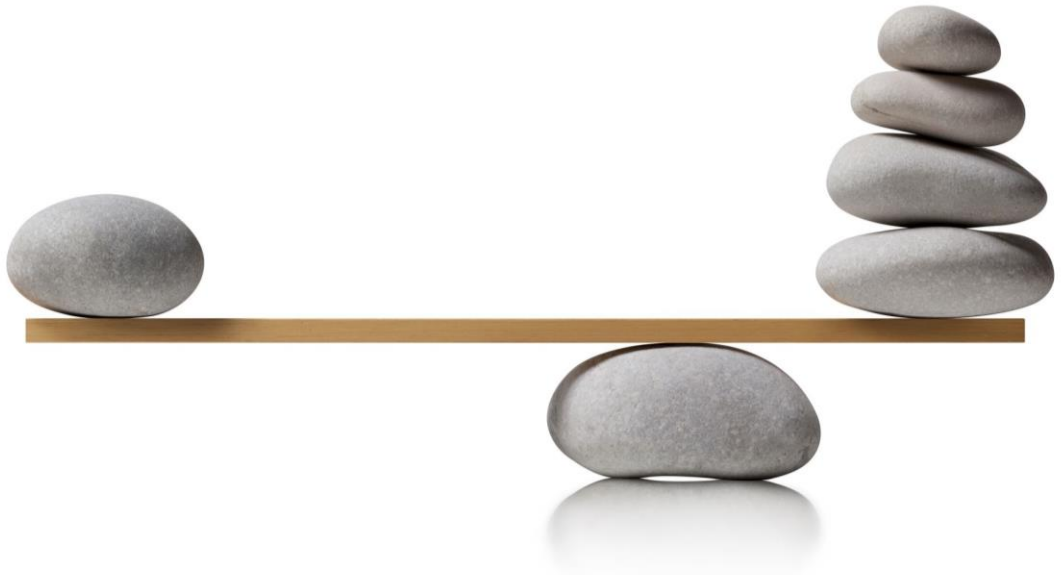
$$P(\neg cancer \mid +) = \frac{0.03 \times 0.992}{P(+) = 0.0376}$$

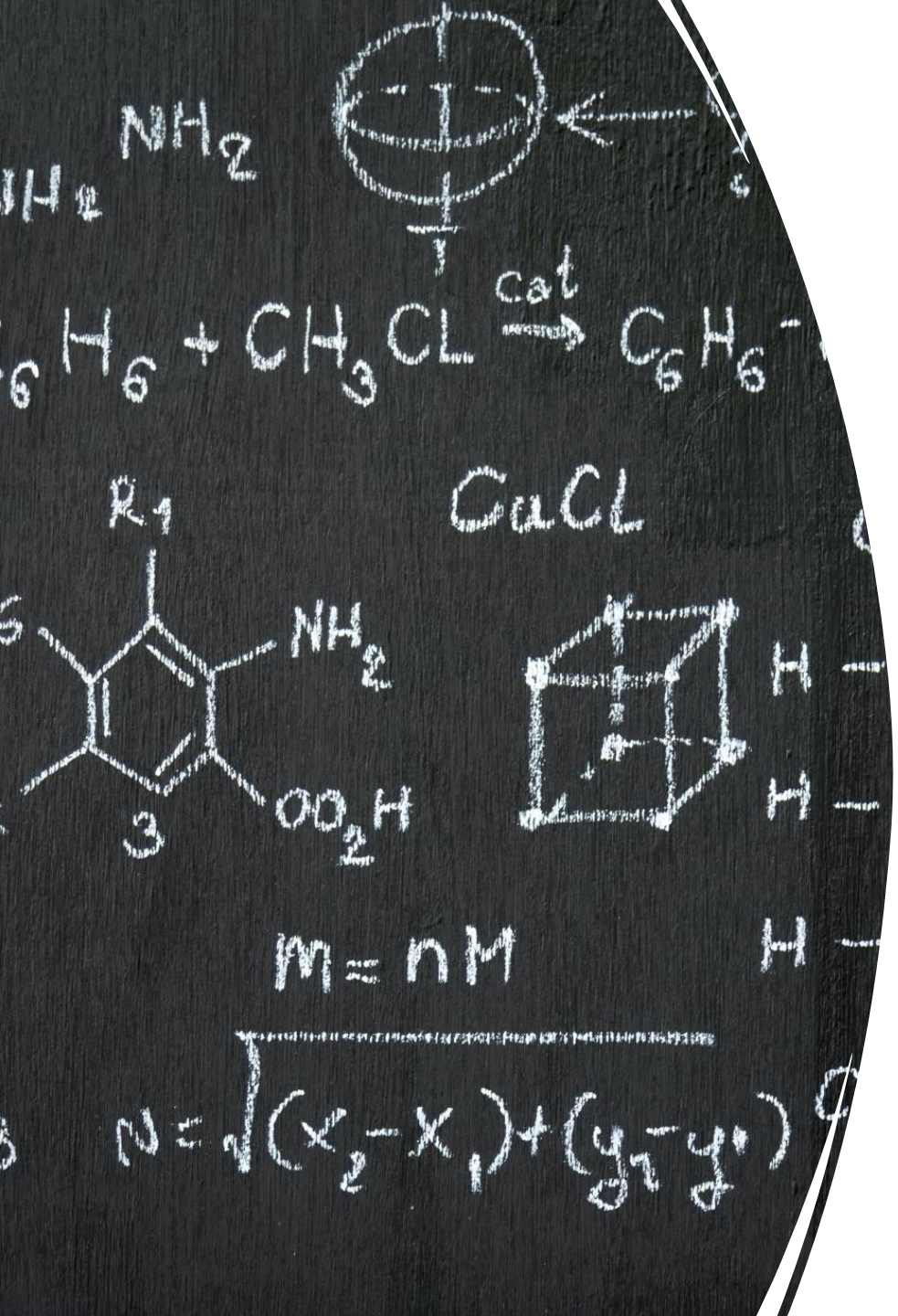
Class Imbalance

$$P(cancer \mid +) = \frac{0.98 \times 0.008}{P(+) = 0.0376} = 0.2085$$

$$P(\neg cancer \mid +) = \frac{0.03 \times 0.992}{P(+) = 0.0376} = 0.7915$$

$$P(cancer \mid +) < P(\neg cancer \mid +)$$





Naïve Bayes Classifier

In this equation, which seems very simple, the term maximum argument, meaning "the one with the greater probability wins", is used.

$$C = \arg\max_{c_j \in C} \left(P(c_j) \prod_{i=1}^{n} P(f_i \mid c_j) \right)$$

Example-2

Find the class of

<Yes, No, Yes, Yes>

f_1	f_2	f_3	f_4	Class
Yes	No	No	Yes	B
Yes	No	No	No	B
No	Yes	Yes	No	M
No	No	Yes	Yes	M
Yes	No	No	Yes	B
Yes	No	No	No	M
Yes	Yes	Yes	No	M
Yes	Yes	No	Yes	M
No	No	No	Yes	B
No	No	Yes	No	M

Example-2

f_1	f_2	f_3	f_4	Class
Yes	No	No	Yes	B
Yes	No	No	No	B
No	Yes	Yes	No	M
No	No	Yes	Yes	M
Yes	No	No	Yes	B
Yes	No	No	No	M
Yes	Yes	Yes	No	M
Yes	Yes	No	Yes	M
No	No	No	Yes	B
No	No	Yes	No	M

$$P(Class = B \mid f_1, f_2, f_3, f_4)$$

$$= \frac{2}{5}\left(\frac{3}{4}1\frac{1}{4}\frac{3}{4}\right) = \frac{9}{160}$$

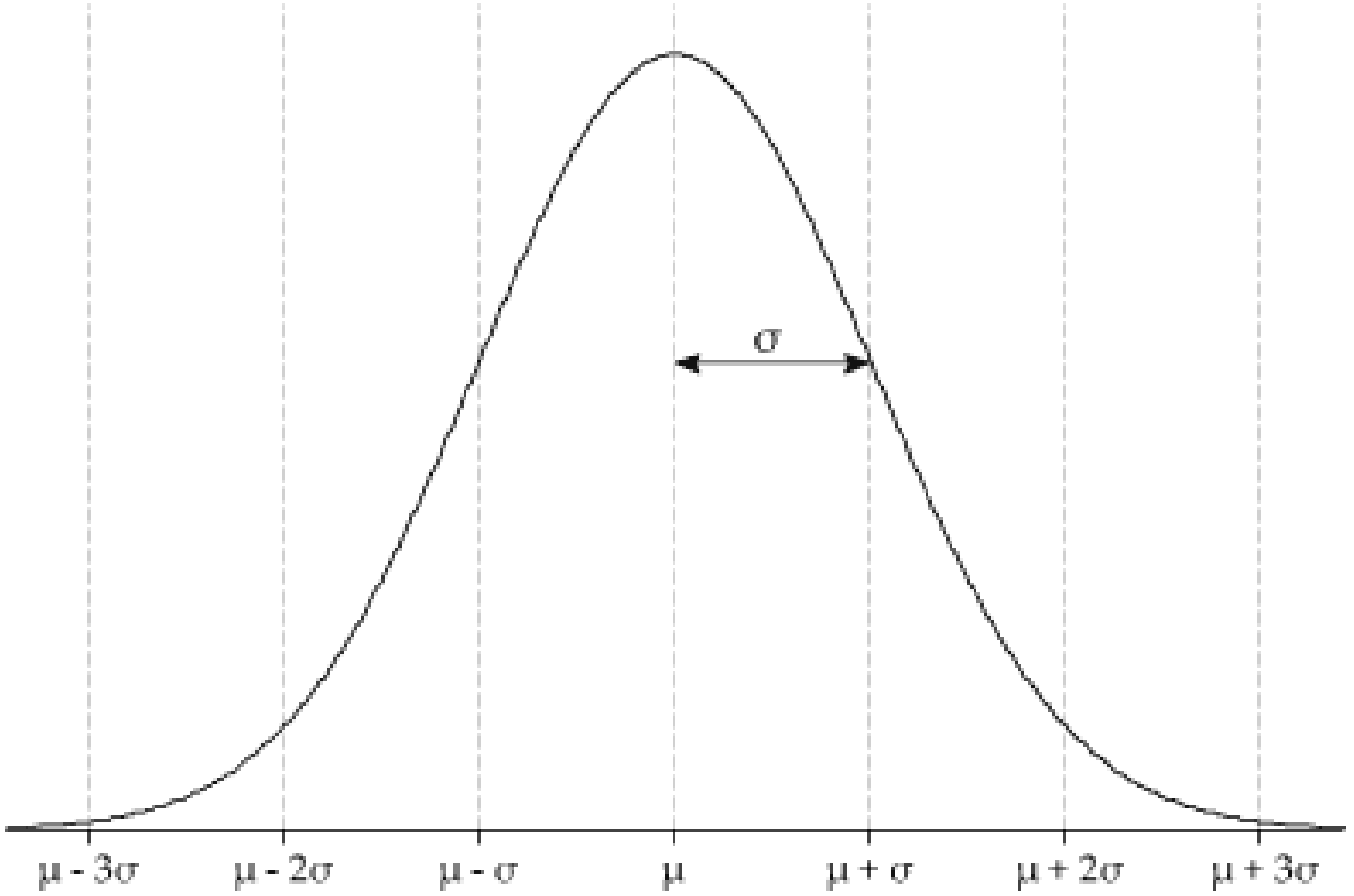
$$P(Class = M \mid f_1, f_2, f_3, f_4)$$

$$= \frac{3}{5}\left(\frac{1}{2}\frac{1}{2}\frac{2}{3}\frac{1}{3}\right) = \frac{1}{30}$$

Numerical Data

Gaussian (Normal) Distribution

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



Example-3

Find the class of

<Yes, Yes, No, No, 45>

V_1	V_2	V_3	V_4	Age	Class
Yes	No	Yes	No	38	F
Yes	Yes	Yes	No	40	F
Yes	Yes	Yes	No	41	F
No	No	No	No	55	F
No	Yes	No	No	27	M
Yes	Yes	Yes	Yes	30	M
Yes	No	Yes	Yes	35	M
No	No	No	No	42	M
Yes	No	No	No	43	M
Yes	No	No	No	45	M

Example-3

For P(Age | F), $\mu_F = 43.5$ $\sigma_F = 7.77$

For P(Age | M), $\mu_M = 37$ $\sigma_M = 7.46$

From Normal distribution,

P(Age=45 | F) = 0.0504

~ 5/100

P(Age=45 | M) = 0.0301

~ 3/100

Age	Class
38	F
40	F
41	F
55	F
27	M
30	M
35	M
42	M
43	M
45	M

Example-3

P (M | Yes, Yes, No, No, 45) = ?

$$P(M|v_1, v_2, v_3, v_4, Age) =$$

$$\frac{3}{5}\left(\frac{2}{3}\frac{1}{3}\frac{2}{3}\frac{2}{3}0.03\right) = 0.0018$$

P (B | Yes, Yes, No, No, 45) = ?

$$P(B|v_1, v_2, v_3, v_4, Age) =$$

$$\frac{2}{5}\left(\frac{3}{4}\frac{1}{2}\frac{1}{4}\frac{4}{4}0.05\right) = 0.0019$$

Class B

Zero Frequency

P(Class=F | V_4 = Yes) = ?

V_4	Class
No	F
No	F
No	F
No	F

Manual Calculations

Find the class of (K, P, T, 6). Let the Laplacian smoothing coefficient be 10^{-2}.

X_1	X_2	X_3	X_4	D
K	P	G	2	H
K	N	G	3	H
B	E	G	1	H
B	P	G	3	S
B	N	T	5	S
K	E	T	4	S

Manual Calculations

Find the class of (K, P, T, 6). Let the Laplacian smoothing coefficient be 10^{-2}.

$\mu_H = 2 \quad \sigma_H = 1 \quad \mu_S = 4 \quad \sigma_S = 1$

$P(X_4 = 6|H) = 0.00013$

$P(X_4 = 6|S) = 0.054$

$$P(H \mid K, P, T, 6) = P(H)\ P(K|H)\ P(P|H)\ P(T|H)\ P(6|H)$$
$$= \frac{1.01}{2.01} * \frac{2.01}{3.01} * \frac{1.01}{3.01} * \frac{0.01}{3.01} * 0.00013$$

$$P(S \mid K, P, T, 6) = P(S)\ P(K|S)\ P(P|S)\ P(T|S)\ P(6|S)$$
$$= \frac{1.01}{2.01} * \frac{1.01}{3.01} * \frac{1.01}{3.01} * \frac{2.01}{3.01} * 0.054$$

Class ***S*** wins

X_1	X_2	X_3	X_4	D
K	P	G	2	H
K	N	G	3	H
B	E	G	1	H
B	P	G	3	S
B	N	T	5	S
K	E	T	4	S

Machine Learning

3. week

Thanks for watching