③

ⓐ The Naive Bayes assumption is that each feature makes an :

① Independent

② Equal

contribution to the outcome.

$$P(y \mid x_1, \ldots, x_n) = \frac{P(x_1|y)P(x_2|y)\ldots P(x_n|y)P(y)}{P(x_1)P(x_2)\ldots P(x_n)}$$

For all entries in the dataset, the denominator does not depend on y and the values of the feature x are given, so that the denominator is effectively constant.

$$y = \text{argmax}_y P(y) \prod_{i=1}^{n} P(x_i|y)$$

Using the above function, we can obtain the class, given the Predictors/Features.

Example:

Weather Dataset

Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Sunny	Mild	High	False	Yes
⋮	⋮	⋮	⋮	⋮

<u>Steps</u>

Posterior Probability $P(y|x)$:

① Calculate the Prior Probability for given class labels using Frequency table for each feature against the target.

② Draw Likelihood Probability from frequency table for each category of feature against the target.

③ Put these values in Bayes formula and calculate Posterior Probability.

④ See which class has higher Probability, given the input belong to the higher Probability class.

ⓒ ⓘ Here Given $P(w_1) = P(w_2) = 1/2$

So, α_1 : Deciding w_1

α_2 : Deciding w_2

0-1 loss function :

	w_1	w_2
α_1	0	1

So,

$P(x|w_1) \sim N(0,1)$ i.e. mean $(\mu) = 0$, $\sigma^2 = 1$

$P(x|w_2) \sim N(1,2)$ i.e. mean $(\mu) = 1$, $\sigma^2 = 2$

So, PDF of evidence i.e. $P(x)$

$$P(x) = P(x|w_1).P(w_1) + P(x|w_2).P(w_2)$$

$$= \left[\frac{1}{\sqrt{2\pi}.\sqrt{1}}.e^{-\frac{(x-0)^2}{2.1}}\right].1/2 + \left[\frac{1}{\sqrt{2\pi}.\sqrt{2}}.e^{-\frac{(x-1)^2}{2.2}}\right]1/2$$

$$\Rightarrow P(x) = \frac{1}{2\sqrt{2\pi}}.e^{\frac{-x^2}{2}} + \frac{1}{2\sqrt{4\pi}}.e^{-\frac{(x-1)^2}{4}}$$

So, $$P\left(\frac{w_1}{x}\right) = \frac{P(x|w_1).P(w_1)}{P(x)}$$

$$P(w_1|x) = \frac{\frac{1}{\sqrt{2\pi}}.e^{-\frac{x^2}{2}}.\frac{1}{2}}{\frac{1}{\sqrt{2\pi}}.e^{-\frac{x^2}{2}}.\frac{1}{2} + \frac{1}{\sqrt{4\pi}}.e^{-\frac{(x-1)^2}{4}}*\frac{1}{2}}$$

$$P(w_1|x) = \frac{\frac{1}{\sqrt{4\pi}}.e^{-\frac{(x-1)^2}{4}}.\frac{1}{2}}{\frac{1}{\sqrt{2\pi}}.e^{-x^2/2}.1/2 + \frac{1}{\sqrt{4\pi}}.e^{-\frac{(x-1)^2}{4}}.\frac{1}{2}}$$

So, let us perform classification by minimizing the risk such that — "decide w_1" else "decide w_2"

$R(\alpha_1|x) < R(\alpha_2|x)$, then

So, $\boxed{P(w_1|x) > P(w_2|x)}$

$$\frac{\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \cdot \frac{1}{2}}{\frac{1}{2\sqrt{2\pi}} \cdot e^{-x^2/2} + \frac{1}{2\sqrt{4\pi}} \cdot e^{-\frac{(x-1)^2}{4}}} > \frac{\frac{1}{\sqrt{2}} \cdot \frac{1}{2\sqrt{4\pi}} \cdot e^{-\frac{(x-1)^2}{4}}}{\frac{1}{2\sqrt{2\pi}} \cdot e^{-x^2/2} + \frac{1}{2\sqrt{4\pi}} \cdot e^{-\frac{(x-1)^2}{4}}}$$

$$\Rightarrow e^{-x^2/2} > \frac{1}{\sqrt{2}} \cdot e^{-(x-1)^2/4}$$

$$\Rightarrow \frac{e^{-x^2/2}}{e^{-\frac{(x-1)^2}{4}}} > \frac{1}{\sqrt{2}}$$

$$e^{\left[-x^2/2 + (x-1)^2/4\right]} > \frac{1}{\sqrt{2}}$$

$$\Rightarrow e^{\left[-x^2/2 + (x-1)^2/4\right]} > \log(1/\sqrt{2})$$

$$\Rightarrow \log\left(e^{\left[-x^2/2 + (x-1)^2/4\right]}\right)$$

$$-x^2/2 + (x-1)^2/4 > \log(1/\sqrt{2})$$

$$\Rightarrow -x^2/2 + x^2 - 2x + 1/4 > \log(1/\sqrt{2})$$

$$\Rightarrow \frac{-2x^2 + x^2 - 2x + 1}{4} > \log(1/\sqrt{2})$$

$$-x^2 - 2x + 1 > 4\log(1/\sqrt{2})$$

$$\Rightarrow x^2 + 2x + 4\log(1/\sqrt{2}) - 1 < 0$$

$$\Rightarrow x^2 + 2x - 2.39 < 0 \rightarrow \text{decide } w_1$$

else

decide w_2

if $-2.84 < x < 0.84 \rightarrow$ Decode w_1

$$x \geq 0.84$$
$$\Rightarrow \boxed{x \leq -2.84} \rightarrow \text{Decode } w_2.$$

(ii) $P(w_1) = 0.6$

$P(w_2) = 0.4$

0-1 Loss function :→

$$g_1(x) \geq g_2(x)$$
$$= P(w_1|x) \geq P(w_2|x)$$
$$\Rightarrow P(x|w_1)\,P(w_1) \geq P(x|w_2).P(w_2)$$
$$\Rightarrow \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-1/2\frac{(x-\mu_1)^2}{\sigma^2}} * 0.6 \geq \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-1/2\frac{(x-\mu_2)^2}{\sigma^2}} * 0.4$$

$$\Rightarrow \frac{-1}{2}\left(\frac{x-0)^2}{1}\right) + \log\left(\frac{0.6}{1}\right) \geq \frac{-1}{2}\left(\frac{(x-1)^2}{2}\right) + \log(0.4/\sqrt{2})$$

$$\Rightarrow -x^2/2 - 0.51 \geq \frac{(-x^2+1-2x)}{4} - 1.26$$

$$\Rightarrow -2x^2 - 2.04 \geq -x^2 - 1 + 2x - 5.04$$

$$\Rightarrow x^2 + 2x - 4 \leq 0$$

$$\boxed{x = 1.23, -3.23}$$

Now by Putting $g(x)$ in equation (i) ⇒

$$P(w_1|x) \geq P(w_2|x)$$

$\Rightarrow P(x|w_1) \cdot P(w_1) \geq P(x|w_2) \cdot P(w_2)$

$$\Rightarrow \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot e^{-\frac{1}{2}\frac{(x-\mu_1)^2}{\sigma^2}} \cdot \cancel{P(w_1)} \geq \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\frac{(x-\mu_2)^2}{\sigma^2}} \cdot \cancel{P(w_2)}$$

$$\Rightarrow \log_e 1^0 - \frac{1}{2}\frac{(x-0)^2}{1} \geq -\frac{1}{2} \cdot \frac{(x-1)^2}{2} + \log_e \frac{1}{\sqrt{2}}$$

$$\Rightarrow -x^2 \geq 2x - 2.38$$

$$\Rightarrow x^2 + 2x - 2.38 \leq 0$$

$$\Rightarrow x = \frac{-2 \pm \sqrt{4 + 4 * 2.38}}{2}$$

$$\boxed{x = 0.84, -2.84}$$

(d) (i) Thus we have if $\sigma_{ij} = 0$ and $\sigma_{ii} = \sigma_i^2$ then

$$\Sigma = \mathrm{diag}(\sigma_1^2, \dots \sigma_d^2)$$

$$= \begin{pmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_d^2 \end{pmatrix}$$

Thus the determinant and inverse matrix are particularly simple.

$$|\Sigma| = \prod_{i=1}^{d} \sigma_i^2$$

$$\Sigma = diag(1/\sigma_1^2, \dots, 1/\sigma_d^2)$$

This leads to the density being expressed as:

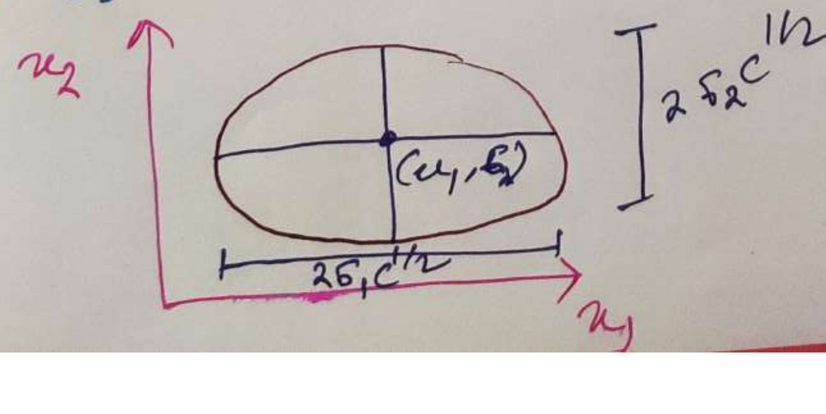
$$P(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \cdot \exp\left[\frac{-1}{2}(x-\mu)^t [diag(1/\sigma_1^2 \dots 1/\sigma_d^2)](x-\mu)\right]$$

$$= \frac{1}{\prod_{i=1}^{d} \sqrt{2\pi}\sigma_i} \cdot \exp\left[\frac{-1}{2}\sum_{i=1}^{d}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2\right]$$

(ii) The contours of constant density are concentric ellipse in d dimensions whose centers are at $(\mu_1, \dots, \mu_d)^t = \mu$, and whose axes in the ith direction are of lengths $2\sigma_i\sqrt{c}$ for the density $P(x)$ held constant at

$$\frac{e^{-c/2}}{\prod_{i=1}^{d}\sqrt{2\pi}\,\sigma_i}$$

The axes of the ellipse are parallel to the co-ordinate axes. The plot in 2 dimensions $(d=2)$ is shown



(b) Answer

(ii) Whether we assume full class covariance matrices or diagonal class covariance matrices.