06/02/24

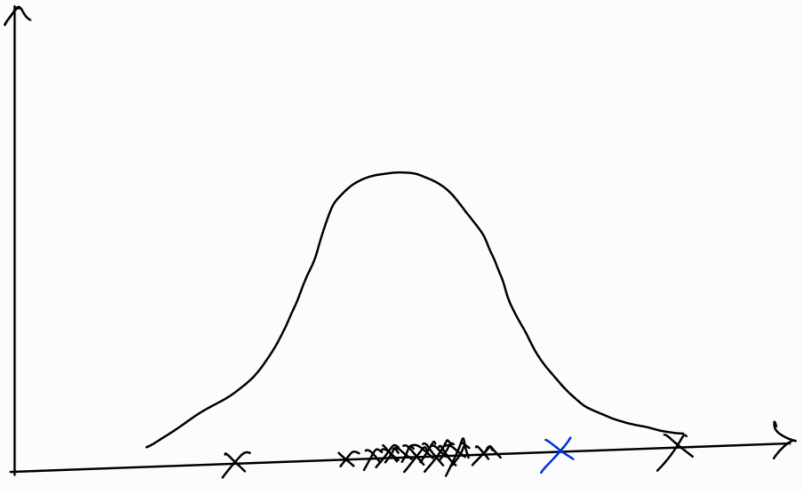
$$E(X) = \sum x \cdot p(x)$$

When we need to sample, we use CDF.

1] Draw CDF CDF = $\int f(x)$ → distribution

2] Select random number

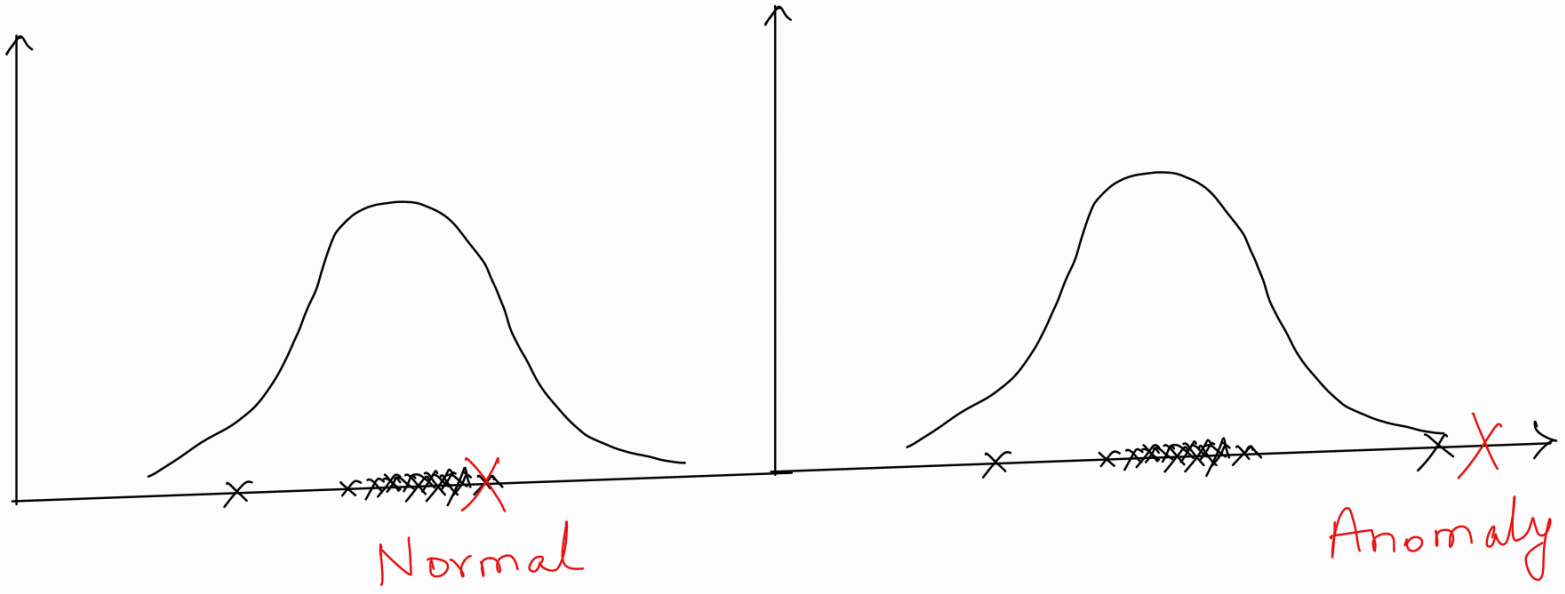
3] Distribution is largely Gaussian



Prob of blue point belonging to PDF.

Put value in Gaussian equation and find likelihood.

→ Anomaly detection is possible using density estimation. If the data point does not lie in the normal range, it can be flagged as anomaly.



$$P(x|y_1, y_2) = P(x|y_1)\,P(x|y_2)$$

Naive Bayes

→ Will not work if data is correlated

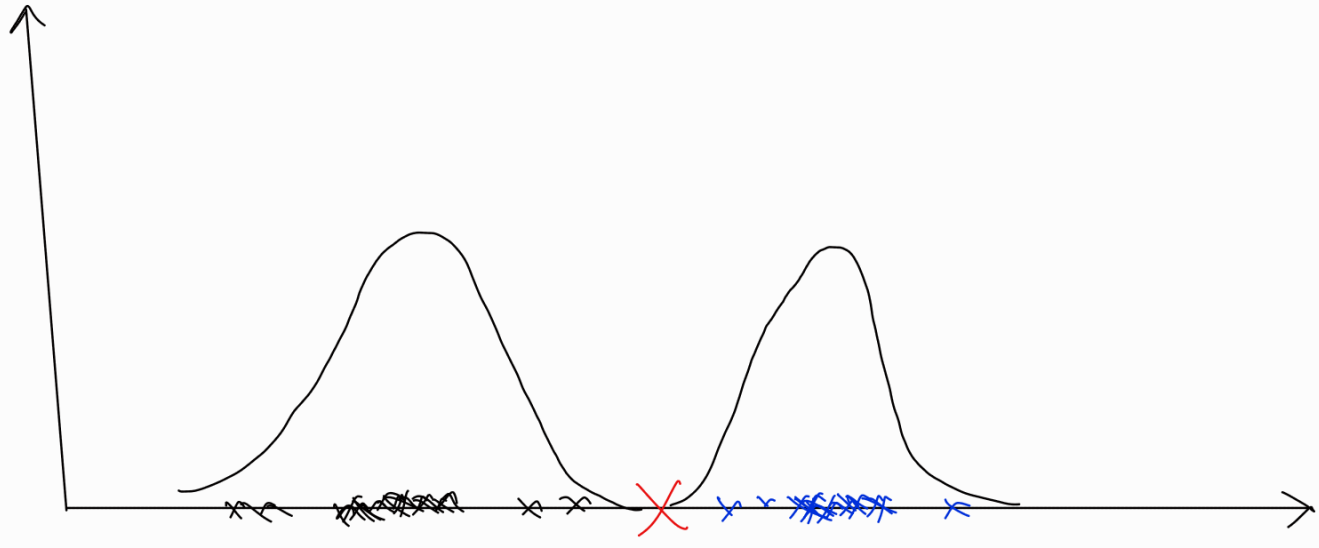
Multivariate Density Estimation

$$f(x) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{\left[\frac{-1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right]}$$

Slide 19

If covariance was 0.4 in middle block, tilt will be visible.

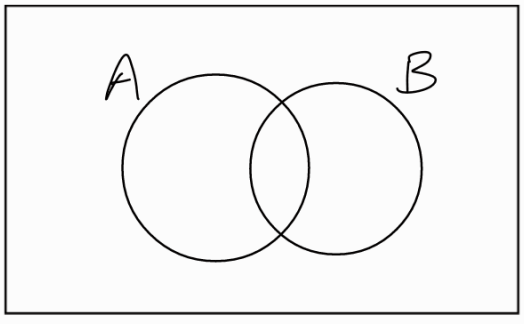
Generative Classification



Compute $P(y_1|x)$ and $P(y_2|x)$

$$\underbrace{P(y_1|x)}_{\text{Posterior}} = \frac{\overbrace{P(x|y_1)}^{\text{likelihood}} \cdot \overbrace{P(y_1)}^{\text{Prior}}}{P(x)}$$

$$P(x) = P(x|y_1)\,P(y_1) + P(x|y_2)\,P(y_2)$$



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

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

Use this to find $P(A)$ and $P(B)$

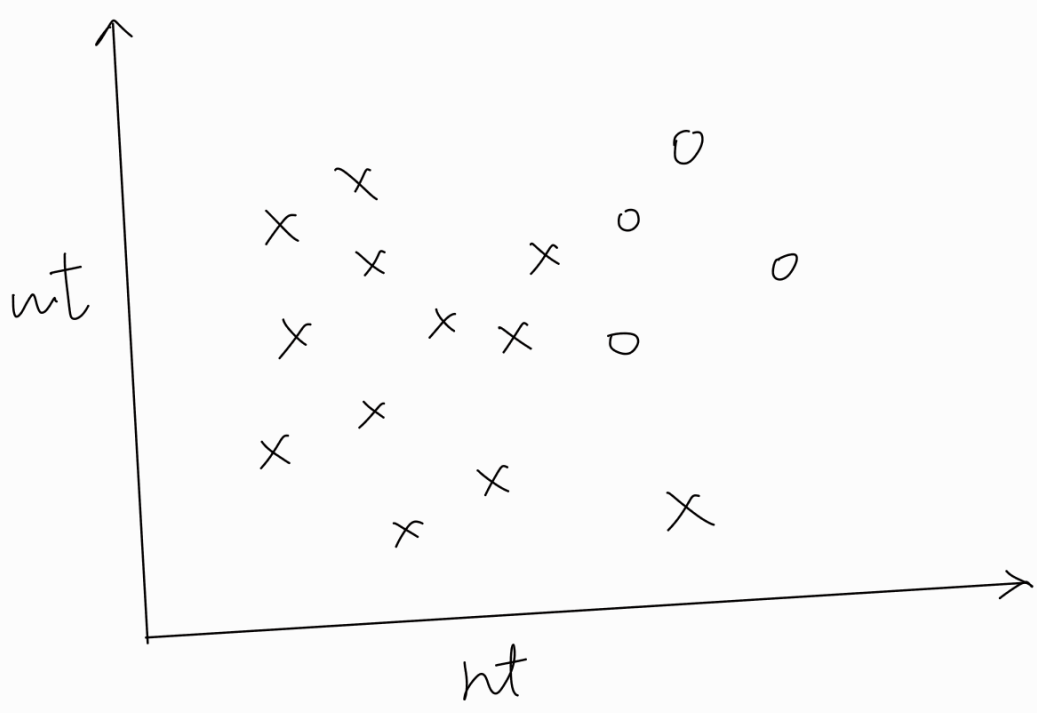
Find $P(y_1|x)$ and $P(y_2|x)$ and assign the class whose probability is higher.

→ Ignore denominator for classification tasks.

Naive Bayes

$$P(x_1, \ldots, x_d | y) = \pi$$

eg]



$$P(a) = \frac{4}{16} \qquad P(c) = \frac{12}{16}$$

$$P(x|a) = P(h_x|a) \cdot P(w_x|a)$$

$$P(x|c) = P(h_x|c) \cdot P(w_x|c)$$

→ This is possible as weight and height are not correlated. Take as independent events & multiply probs.

Spam Classification

→ Data follows multinomial distribution

Dictionary : Dear, Friend, Lunch, Money

$P(N) = 0.67$

$P(\text{Dear} \mid N) = 0.471$

$P(\text{Friend} \mid N) = 0.291$

$P(\text{Lunch} \mid N) = 0.181$

$P(\text{Money} \mid N) = 0.061$

$P(S) = 0.33$

$P(\text{Dear} \mid S) = 0.29$

$P(\text{Friend} \mid S) = 0.14$

$P(\text{Lunch} \mid S) = 0$

$P(\text{Money} \mid S) = 0.57$

Email : Dear friend

$$P(N \mid \text{"Dear Friend"}) = P(\text{"D.F"} \mid N) \cdot P(N)$$

$$= P(\text{Dear} \mid N)\, P(\text{Friend} \mid N) \cdot P(N)$$

$$= 0.471 \times 0.291 \times 0.67$$

$$= 0.091$$

$$P(S \mid \text{"DF"}) = 0.29 \times 0.14 \times 0.33$$

$$= 0.013$$

∴ The email classified as normal.

Email: Lunch Money Money Money

$$P(N \mid \text{Email}) = (0.181) \times (0.061)^3 \times 0.67 = 0.00002$$

$$P(S \mid \text{Email}) = \underbrace{0.001} \times (0.57)^3 \times 0.33 = 0.0006$$

$P(\text{Lunch} \mid S) = 0$. But we add 0.001 to every probability.

∴ Email classified as spam.