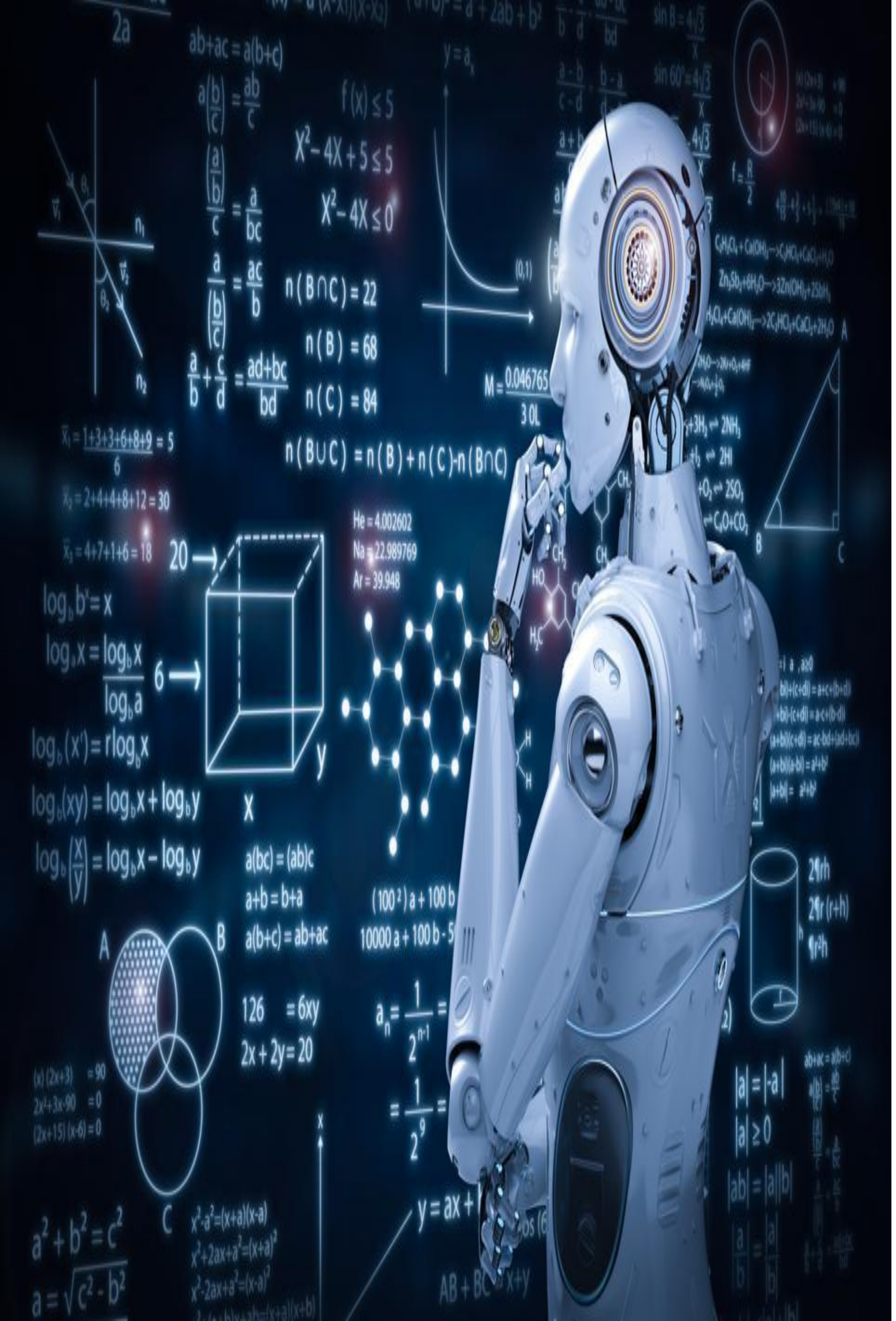


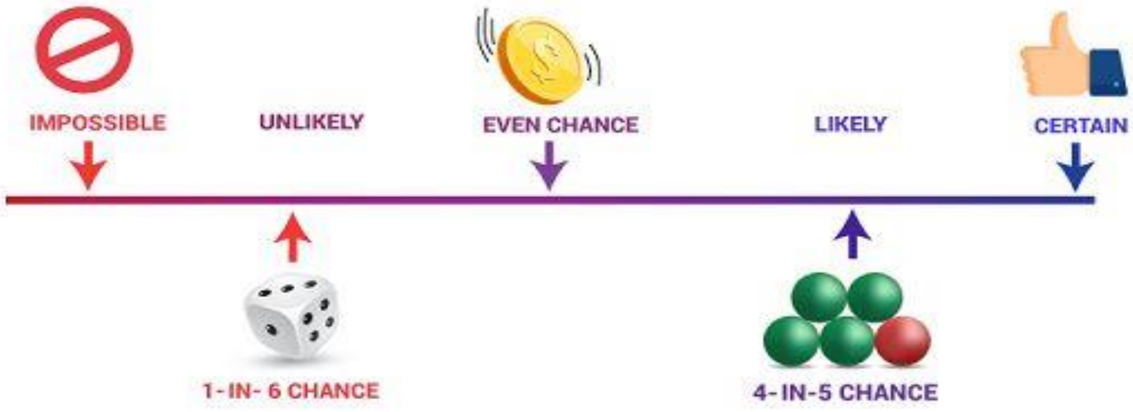
Practical Machine Learning

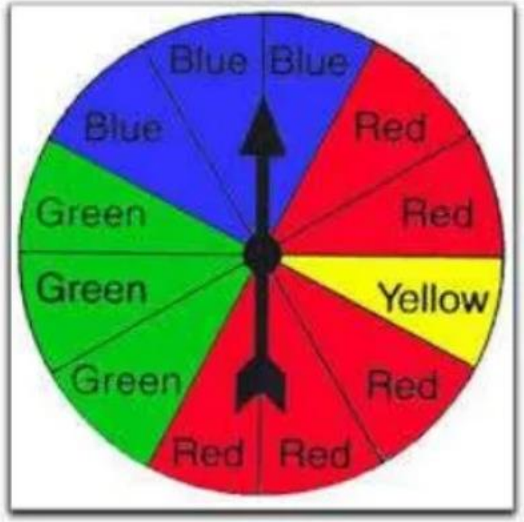
Day 10: Mar23 DBDA

Kiran Waghmare

Agenda

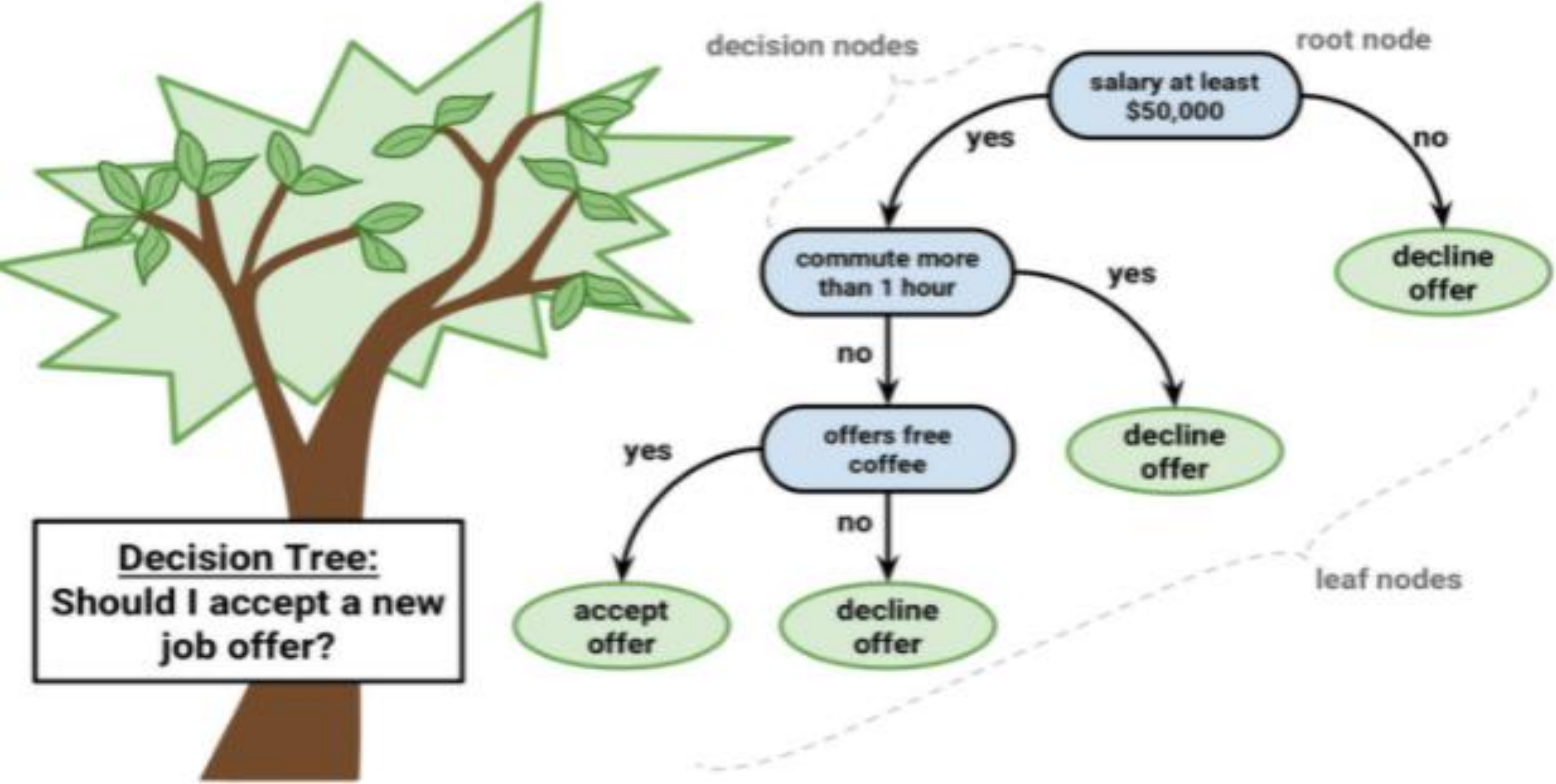
- Naïve Bayes
- Decision Tree





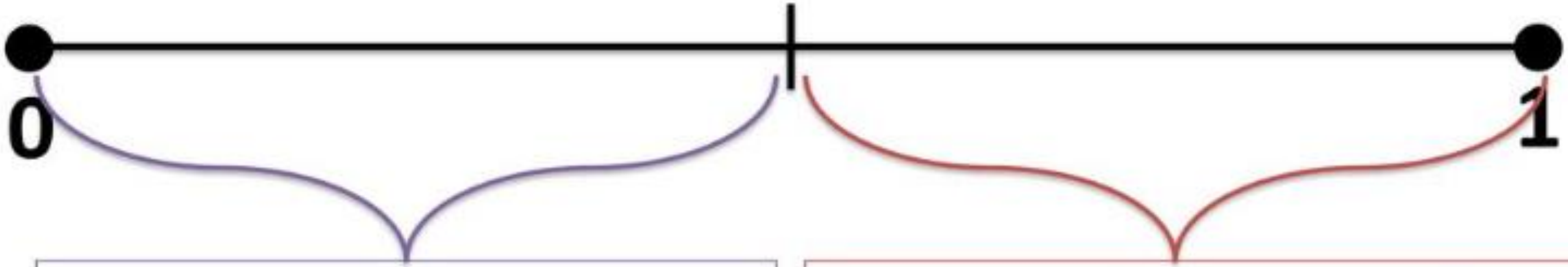
PROBABILITY

GETTING KNOWLEDGE READY


Decision Tree:
Should I accept a new job offer?
decision nodes
root node
salary at least $50,000
yes
no
commute more than 1 hour
decline offer
no
yes
offers free coffee
decline offer
yes
no
accept offer
decline offer
leaf nodes

THE PROBABILITY IN EVERY LIFE

Probability Number Line



An event whose probability is closer to 0 is less likely to occur.

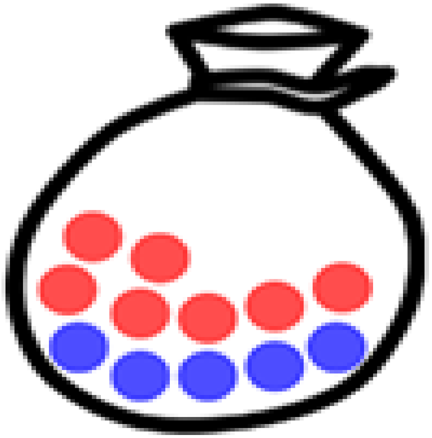
An event whose probability is closer to 1 is more likely to occur.

$$p(head) = 1/2$$

Simple Probability

$$\text{Probabilty} = \frac{\text{Favorable outcomes}}{\text{Total outcomes}}$$

Example:



$$P(red) = \frac{7}{12}$$

Number of red marbles

Total number of marbles (sample space)

$$P(blue) = \frac{5}{12}$$

Number of blue marbles

Total number of marbles (sample space)

Pick a random card, what is the probability of getting a queen?

4 queens, 52 total cards

Pick a random card, you know it is a diamond. Now what is the probability of that card being a queen?

Total diamonds = 13

Queen = 1

Conditional Probability Formula

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

Probability of A given B

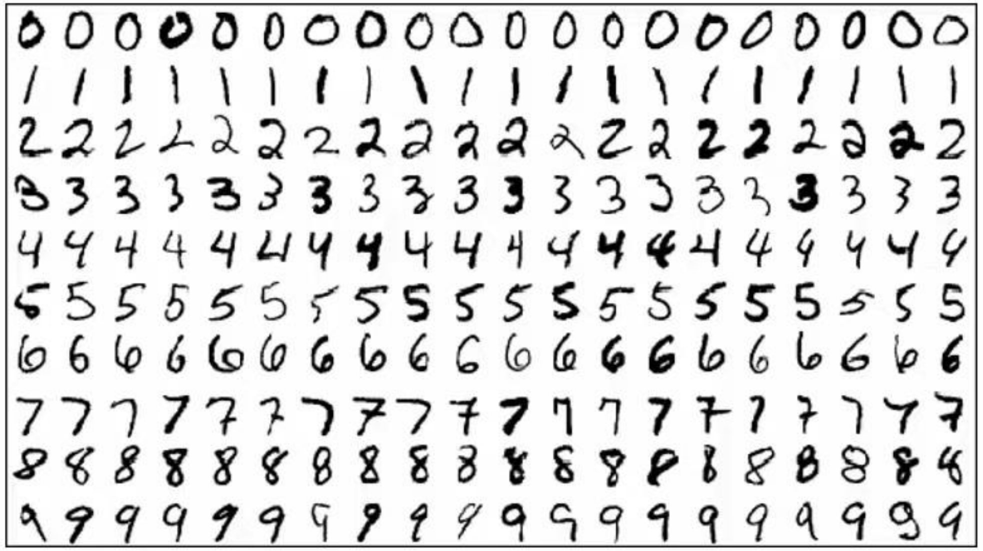
Probability of A and B

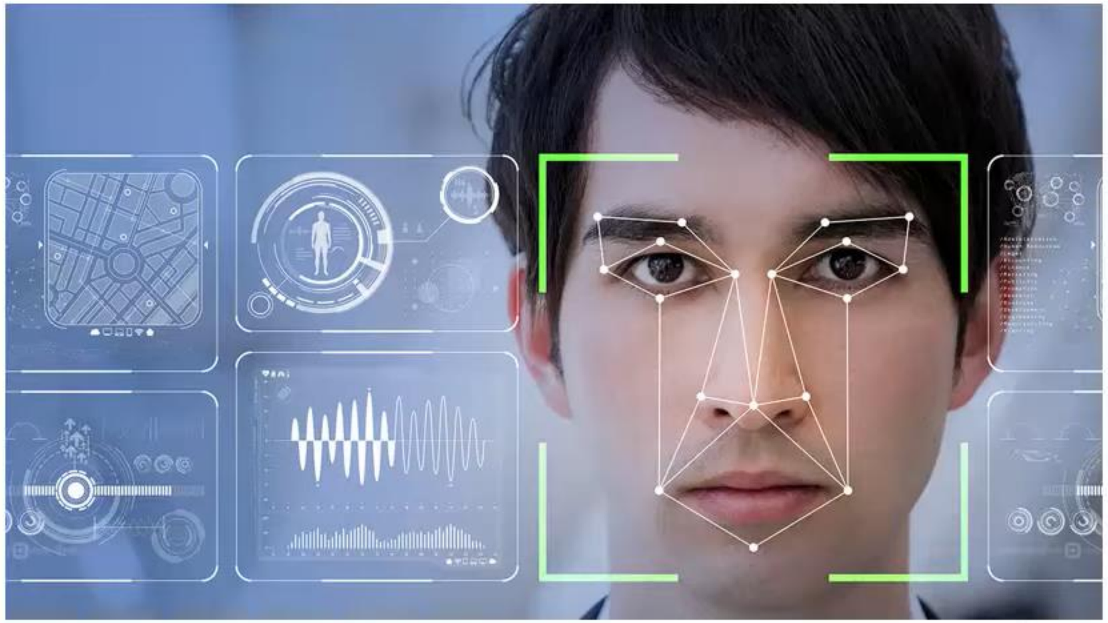
Probability of B

Thomas Bayes

$$P(A/B) = \frac{P(B/A) * P(A)}{P(B)}$$


EMAIL
SPAM
SPAM






NEWS

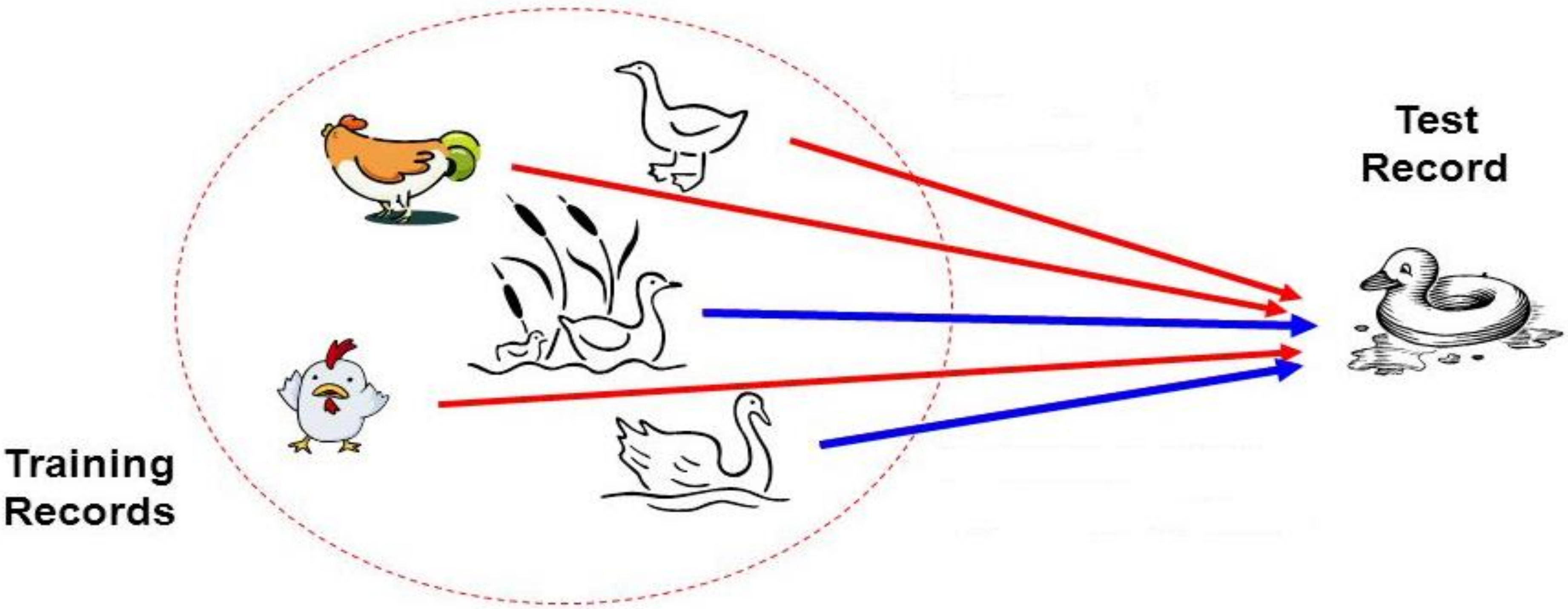
Examples of Classification in Data Analytics

- Life Science: Predicting tumor cells as benign or malignant
- Security: Classifying credit card transactions as legitimate or fraudulent
- Prediction: Weather, voting, political dynamics, etc.
- Entertainment: Categorizing news stories as finance, weather, entertainment, sports, etc.
- Social media: Identifying the current trend and future growth

Bayesian Classifier

Bayesian Classifier

- Principle
  - If it walks like a duck, quacks like a duck, then it is probably a duck



Bayesian Classifier

- **A statistical classifier**
  - Performs *probabilistic prediction*, *i.e.,* predicts class membership probabilities
- **Foundation**
  - Based on Bayes’ Theorem.
- **Assumptions**
  1. The classes are mutually exclusive and exhaustive.
  2. The attributes are independent given the class.
- **Called “Naïve” classifier because of these assumptions.**
  - Empirically proven to be useful.
  - Scales very well.

Whether	Play
Sunny	No
Sunny	No
Overcast	Yes
Rainy	Yes
Rainy	Yes
Rainy	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rainy	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rainy	No

BAYES THEOREM

- Bayes theorem is the cornerstone of Bayesian learning methods because it provides a way to calculate the posterior probability **P(h|D), from**
- **the prior** probability **P(h),**
- **Probability over the data set P(D) and**
- **Current probability P(D(h)**

$$P(h|D) = \frac{P(D|h)p(h)}{P(D)}$$

Maximum A Posteriori (MAP) Hypothesis

- The learner considers some set of candidate hypotheses H and is interested in finding the most probable hypothesis **h ϵ H** given the observed data **D (or at least one of the maximally probable if there** are several).
- Any such maximally probable hypothesis is called a **maximum a posteriori (MAP) hypothesis.**
- **We can determine the MAP hypotheses by using** Bayes theorem to calculate the posterior probability of each candidate hypothesis.

Maximum A Posteriori (MAP) Hypothesis

- More precisely, we will say that h_{MAP} **is a MAP hypothesis provided**

$$h_{MAP} \equiv \underset{h \in H}{\operatorname{argmax}}\, P(h|D)$$

$$= \underset{h \in H}{\operatorname{argmax}}\, \frac{P(D|h)\,P(h)}{P(D)}$$

$$= \underset{h \in H}{\operatorname{argmax}}\, P(D|h)\,P(h)$$

Whether	Play
Sunny	No
Sunny	No
Overcast	Yes
Rainy	Yes
Rainy	Yes
Rainy	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rainy	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rainy	No

Multnomial Naive Bayes: Example

	docID	words in document	in c = China?
Training set	1	Chinese Beijing Chinese	yes
	2	Chinese Chinese Shanghai	yes
	3	Chinese Macao	yes
	4	Tokyo Japan Chinese	no
Test set	5	Chinese Chinese Chinese Tokyo Japan	?

Example

Given all the previous patients I've seen (below are their symptoms and diagnosis)...

chills	runny nose	headache	fever	flu?
Y	N	Mild	Y	N
Y	Y	No	N	Y
Y	N	Strong	Y	Y
N	Y	Mild	Y	Y
N	N	No	N	N
N	Y	Strong	Y	Y
N	Y	Strong	N	N
Y	Y	Mild	Y	Y

Do I believe that a patient with the following symptoms has the flu?

chills	runny nose	headache	fever	flu?
Y	N	Mild	Y	?

NAIVE BAYES CLASSIFIER – Example -1

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

$$\langle Outlook = sunny, Temperature = cool, Humidity = high, Wind = strong \rangle$$

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

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

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

NAIVE BAYES CLASSIFIER

Example - 1

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

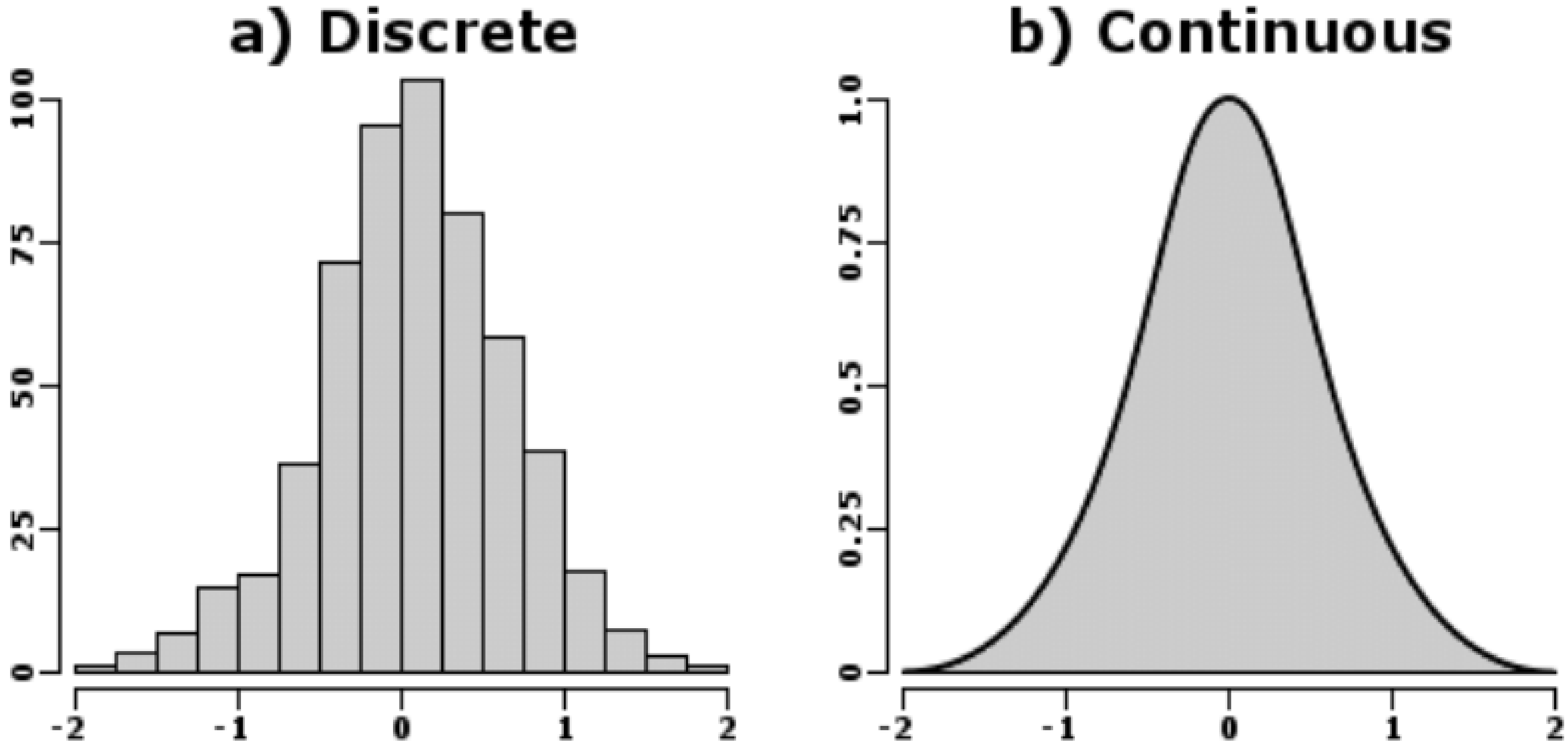
Example - 1

Conclusions

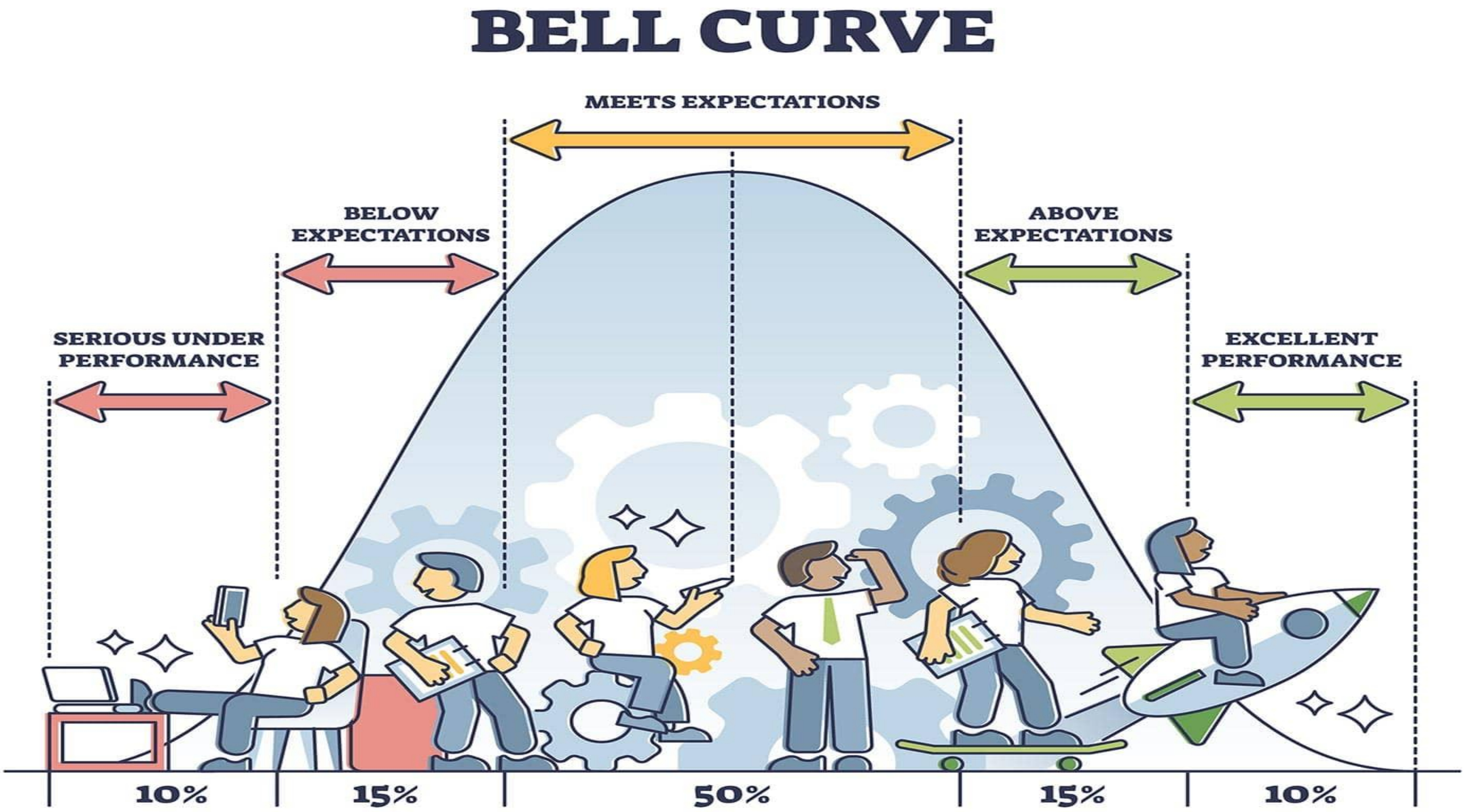
- **Naïve Bayes based on the independence assumption**
  - Training is **very easy and fast**; just requiring considering each attribute in each class separately
  - Test is **straightforward**; just looking up **tables or calculating conditional probabilities with normal distributions**
- **A popular generative model**
  - **Performance competitive** to most of state-of-the-art classifiers even in presence of violating independence assumption
  - Many successful applications, e.g., spam mail filtering
  - Apart from classification, naïve Bayes can do more…

	Discrete		Continuous
Probability Distribution	x_1 x_2 ... x_n	p_1 p_2 ... p_n	pdf: $f(x)$
F(x)	$\sum_{i=1}^{n} p_i = 1$		$\int_{-\infty}^{\infty} f(x)\,dx = 1$
Mean μ	$\sum_{i=1}^{n} x_i p_i$		$\int_{-\infty}^{\infty} x f(x)\,dx$
Variance σ^2	$\sum_{i=1}^{n} (x_i - \mu)^2 p_i$		$\int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$

X	*X Counts*	*p(x)*	*Values of X*	*E(x)*	*V(x)*
Discrete uniform	Outcomes that are equally likely (finite)	$\frac{1}{b-a+1}$	$a \le x \le b$	$\frac{b+a}{2}$	$\frac{(b-a+2)(b-a)}{12}$
Binomial	Number of sucesses in n fixed trials	$\binom{n}{x} p^x (1-p)^{n-x}$	$x = 0, 1, \ldots, n$	np	$np(1-p)$
Poisson	Number of arrivals in a fixed time period	$\frac{e^{-\lambda}\lambda^x}{x!}$	$x = 0, 1, 2, \ldots$	λ	λ
Geometric	Number of trials up through 1st success	$(1-p)^{x-1}p$	$x = 1, 2, 3, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative Binomial	Number of trials up through kth success	$\binom{x-1}{k-1}(1-p)^{x-k}p^k$	$x = k, k+1, \ldots$	$\frac{k}{p}$	$\frac{k(1-p)}{p^2}$
Hyper - geometric	Number of marked individuals in sample taken without replacement	$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$	$\max(0, M+n-N) \le x \le \min(M, n)$	$n * \frac{M}{N}$	$\frac{nM(N-M)(N-n)}{N^2(N-1)}$



Discrete Vs Continuous


BELL CURVE
MEETS EXPECTATIONS
BELOW EXPECTATIONS
ABOVE EXPECTATIONS
SERIOUS UNDER PERFORMANCE
EXCELLENT PERFORMANCE
10%
15%
50%
15%
10%