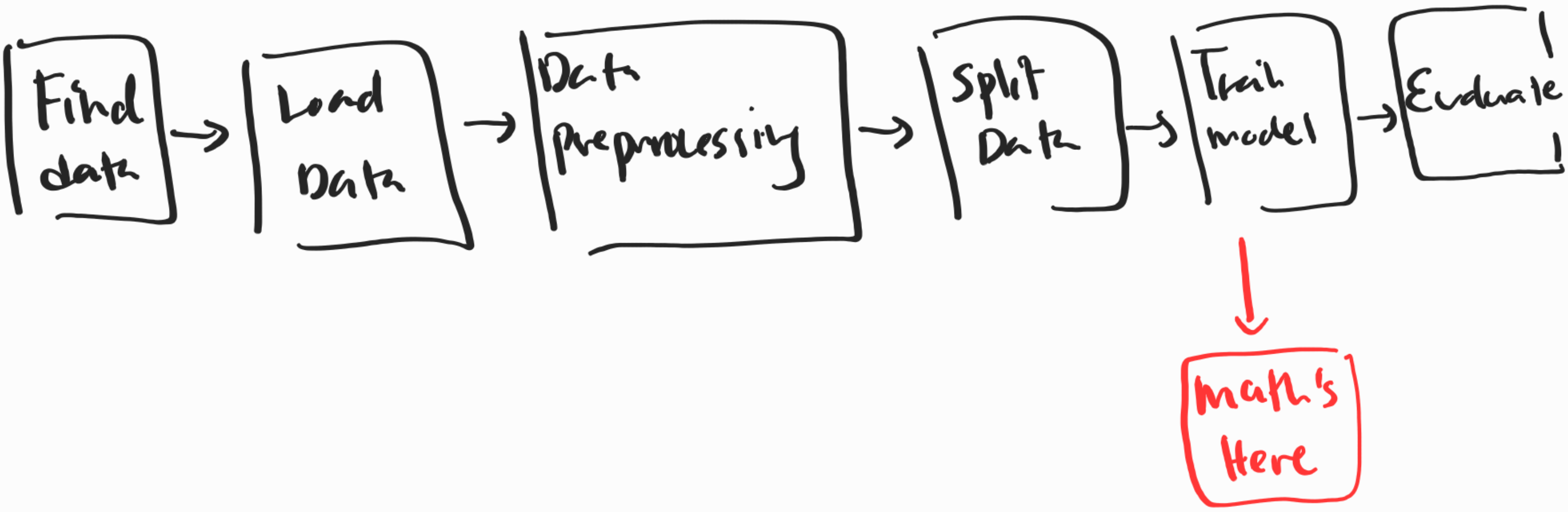
Machine Learning Flow (works for any algo)



Naive Bayes: The Maths

Consider this situation. You have a friend, named Ali. He comes to class. Sometimes he comes by car, sometimes by bus. You've noticed that when he comes by car, he is usually early. When he comes by bus, he is often late.

$P(\text{Bus}) = 0.4$: Probability of Ali taking Bus

$P(\text{Car}) = 0.6$: " " " " Car

↳ Prior probability.

$P(\text{Early} \mid \text{Bus}) = 0.25$ ← Conditional probability

$P(x \mid \text{condition})$

The probability of Ali coming to class Early given that we know he took the Bus.

$P(\text{Marry supermodel})$ ↓ 0.01

$P(\text{Marry supermodel} \mid \text{Rich like oil tycoon}) = 0.4$

Given information

$P(Bus) = 0.4$ $P(Car) = 0.6$

$P(Early \mid Bus) = 0.25$

$P(Late \mid Bus) = 0.75$

$P(Early \mid car) = 0.75$

$P(Late \mid car) = 0.25$

Intuitive, common sense, logic

Joint probability

$P(X,Y) = P(X \mid Y)\, P(Y)$ or $P(Y \mid X)\, P(X)$

The probability of 2 events happening

Marginal Probability

$$P(X) = \sum_Y P(X,Y)$$
$$= \sum P(X \mid Y)\, P(Y)$$

Continue with our example

Early

- $P(Early, Bus)$
  $= P(Early \mid Bus) \cdot P(Bus) =$
  $= (0.25)(0.4) = 0.1$
- $P(Early, car)$
  $= P(Early \mid car)\, P(car) = 0.75 \times 0.6$
  $= 0.45$
- $P(Early) = P(Early, Bus) + P(Early, car)$
  $= 0.1 + 0.45 = 0.55$

Late

$P(Late, Bus)$

$= P(Late \mid Bus)\, P(Bus) = (0.75)(0.4)$

$= 0.3$ #

$P(Late, car)$

$= P(Late \mid car)\, P(car)$

$= (0.25)(0.6) = 0.15$

Let's put everything together

	Bus	Car	
Early	0.1	0.45	0.55
Late	0.3	0.15	0.45
Total	0.4	0.6	1.0

$P(Late, Bus)$

Summary

- ✓ Prior Probability : Given earlier
- ✓ Conditional " :
- ✓ Joint Probability
- ✓ Marginal Probability

P(Bus) = 0.4 P(Car) = 0.6

	Bus	Car	
Early	0.1	0.45	0.55 P(Early)
Late	0.3	0.15	0.45 P(Late)
	0.4 P(Bus)	0.6 P(Car)	1.0

Prior : The probability that is given earlier. Or something that we know beforehand

Marginal : P(Early) : Probability of Ali being early regardless of the mode transportation

Back to ML

What does the dataset looks like?

Day	Bus	Car	Early/Late
1	YES	No	Early
2	YES	No	Late
3	No	YES	Early
⋮	⋮	⋮	⋮
365			

Attributes/Features (Bus, Car) — Target output (Early/Late) — Samples/Observations (rows)

- From the dataset above, you can calculate the conditional probabilities → Refer slides on how to calculate from a table.

  $P(\text{Early} \mid \text{Bus})$ $\quad$ $P(\text{Late} \mid \text{Bus})$

  $P(\text{Early} \mid \text{Car})$ $\quad$ $P(\text{Late} \mid \text{Car})$

- The ML Question $\boxed{\text{CLASSIFICATION}}$

  Given a new row of information like below, what is the prediction for Ali?

Day	Bus	Car	Early/Late
367	YES	No	? Late

Predict

- Posterior Probability, Bayes Theorem

$$P(\text{Bus} \mid \text{Late}) = \frac{P(\text{Late} \mid \text{Bus}) \cdot P(\text{Bus})}{P(\text{Late})}$$

Posterior Probability

- $P(Bus|Late) = \dfrac{P(Late|Bus)P(Bus)}{P(Late)}$

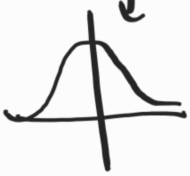
- $P(Bus|Early) = \dfrac{P(Early|Bus)P(Bus)}{P(Early)}$

- $P(Car|Late)$

- $P(Car|Early)$

When we look at the slides, it will tell you, in practical we use the Gaussian Probability

$$\underbrace{P(Bus|Late)}_{\text{Posterior}} = \frac{\overbrace{P(Late|Bus)}^{\text{Likelihood}}\ \overbrace{P(Bus)}^{\text{Prior}}}{\underbrace{P(Late)}_{\text{Marginal}}}$$

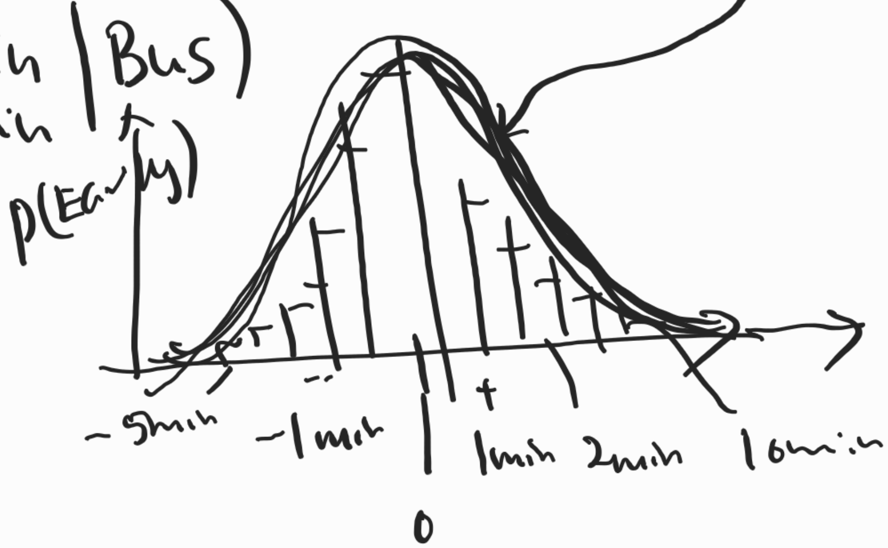


Both the posteriors and likelihood are assumed to be normally distributed.

$$p(x|y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left[\frac{(x-\mu)^2}{2\sigma}\right]}$$

- Example

$P(Early = +1min, +2min \mid Bus)$



Day	Bus	Car	Early (By minuts
Day 1	1	0	+1 min
2	0	1	+2 min
⋮	⋮	⋮	+0
⋮	⋮	⋮	−1 min
⋮	⋮	⋮	−10 mins
365	⋮		

P(Early | Bus)

$$P(\text{Early} \mid \text{Bus}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left[\frac{-(\text{Early}-\mu)^2}{2\sigma}\right]}$$