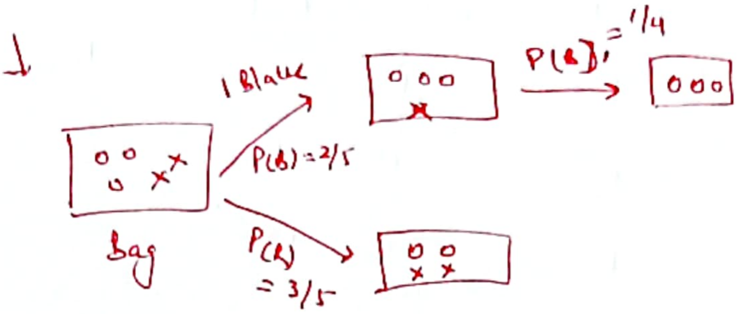
Naive Bayes

- Independent events — Tossing two coins - both outputs are independent to each other
- Dependent events



Conditional Probability = $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

(Probability of A given B)

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

Probability of A

Prob of B given A

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

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

$P(A \cap B) = P(A) \times P(B/A)$

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

Prior Prob.

Likelihood Prob

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

Bayes Theorem

marginal Prob

posterior prob.

Naive Bayes classifier w.r.t text data

f_1 → vectors	f_2 → vectors	f_3	f_4	o/p → +ve / -ve
The	food	delicious	bad	
1	1	1	0	1
1	1	0	1	0
0	1	0	1	0
0	1	1	0	1
0	0	0	1	0

sentence → good / bad → classify

NLP → preprocessing: remove stopwords, stemming, bag of words, TFIDF / words to vec.

~~Here~~ $P(y = yes \mid \text{sentences}) = P(y = yes \mid (x_1, x_2, \ldots x_n))$

$[x_1, x_2, \ldots]$ many words

$$= \frac{P(y) \times \prod_{i=1}^{n} P(x_i \mid y = yes)}{\prod_{i=1}^{n} P(x_i)}$$

multiply.

	x_1	x_2		x_3
sentence - 1 -	The	food	is	delicious
sentence - 2 -	The	food	is	bad
sentence - 3	Food	is	bad	

$P(y = yes) = 2/5$, food, delicious

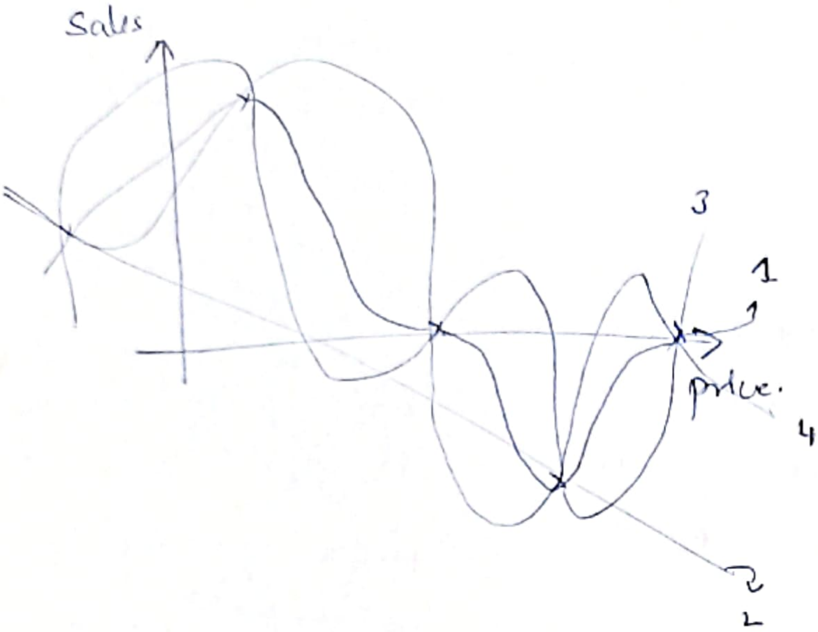
$$= \frac{\cancel{2/5} \times 1/\cancel{8} \times \cancel{3/5} \times 3/5}{\cancel{2/5} \times \cancel{4/5} \times \cancel{3/5}}$$

when the 2 output is 1

$$= \frac{1}{10}$$

(ii) $P(y = no \mid \text{sentence})$ → compare both probability & then decide whether yes or no.

Gaussian Regression

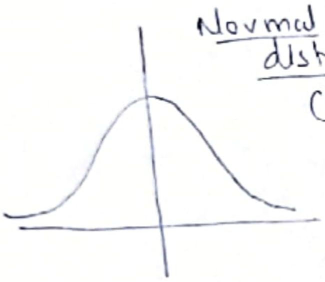


Consider all alternative fits t the points. Consider all functions and error probability and then figure out which is a more likely

GP prediction is a distribution not a concrete number

Gaussian Process

Normal distribution (2D) ⊂ Multivariate Normal distribution (3D) → Infinite vectors (gaussian process)



→ Conditional probability

$$p(f|y_i) = \frac{p(f, y_i)}{p(y_i)}$$

$p(f, y_i)$ → probability of f^n at y_i

$p(y_i)$ → probability of y_i

$p(f|y_i)$ → probability of a f^n given at point y_i

Now to adapt the distribution t new points.

I will now take gaussian distribution + conditional probability t find the best fit.

But we need a guess → one good guess is to take the mean of all of the function. but we may not only have one mean but a σ distribution t define the range of functions.