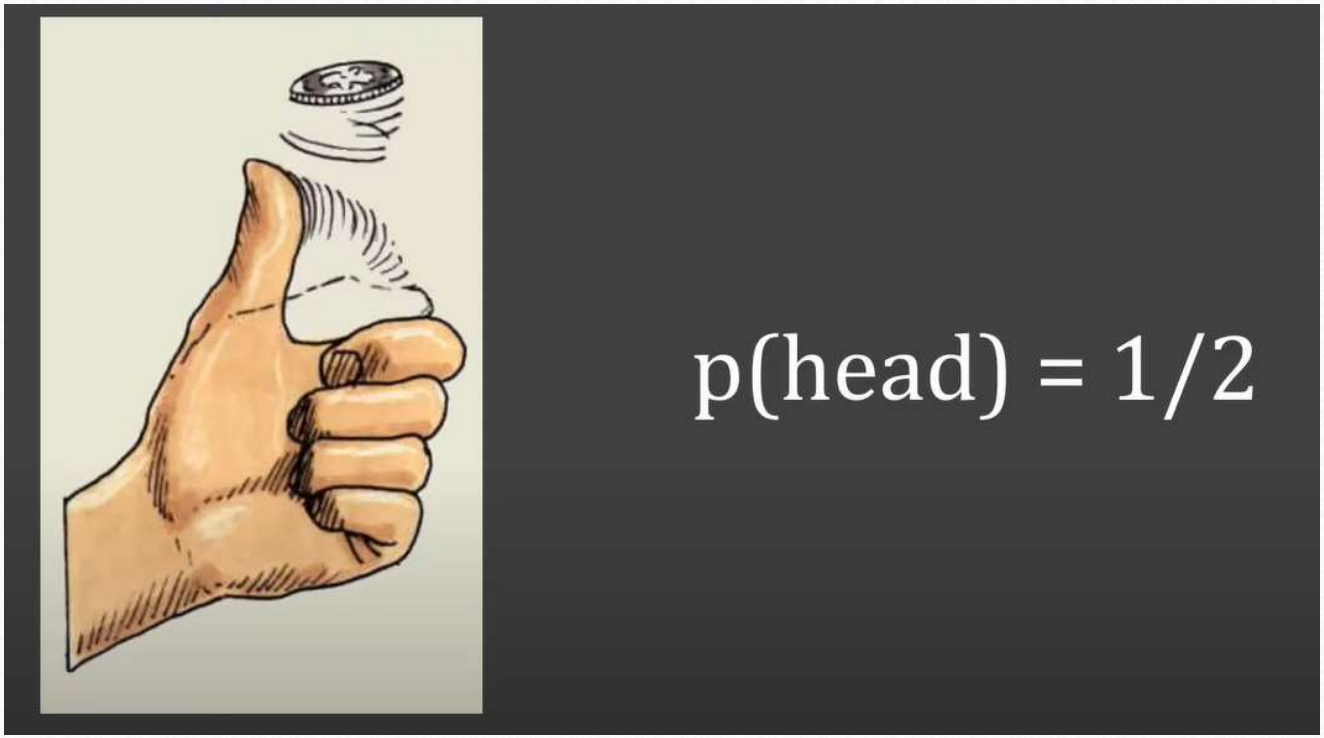
NAÏVE BAYES CLASSIFIER

Objective

After attending this session, you should be able

- Basic of probability, conditional probability
- Introduction of Naïve bayes
- Advantages and disadvantages of Naïve bayes
- Applications
- Hands on exercises

Probability: Flipping the coin



Probability of getting the queen

4 queens, 52 total cards

P(queen) = 4/52 = 1/13

Conditional probability

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

Naïve Bayes

- It is a classification technique based on Bayes' Theorem
- Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.
- Naive Bayes model is easy to build
- Useful for very large data sets.
- Outperform even highly sophisticated classification methods.

Why naive

- Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability of an object and that is why it is known as ‘Naive’.

Conditional Probability

P(queen/diamond)= 1/13

P(A/B)= Probability of event A knowing that event B has already occurred

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

$$P(queen/diamond) = \frac{P(diamond/queen) * P(queen)}{P(diamond)}$$

$P(diamond/queen) = 1/4$

$P(queen) = 1/13$

$P(diamond) = 1/4$

$$= \frac{1/4 * 1/13}{1/4}$$

$$= 1/13$$

Advantages of naïve bayes

- It is easy and fast to predict class of test data set.
- It also perform well in multi class prediction
- Naive Bayes classifier performs better compare to other models like logistic regression
- Require less training data.
- It perform well in case of categorical input variables compared to numerical variables

Disadvantages

- Zero Frequency
- bad estimator.
- Another limitation of Naive Bayes is the assumption of independent predictors.

Applications of Naive Bayes Algorithms

- Real time Prediction
- Text classification
- Spam Filtering
- Sentiment Analysis
- Multi class Prediction
- Recommendation System

Hands on Exercises

Passenger Id	Name	Pclass	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Survived
1	Braund, Mr. Owen Harris	3	male	22	1	0	21171	7.25		S	0
2	Cumings, Mrs. John Bradley	1	female	38	1	0	17599	71.2833	C85	C	1
3	Heikkinen, Miss. Laina	3	female	26	0	0	3101282	7.925		S	1
4	Futrelle, Mrs. Jacques Heath	1	female	35	1	0	113803	53.1	C123	S	1
5	Allen, Mr. William Henry	3	male	35	0	0	373450	8.05		S	0
6	Moran, Mr. James	3	male		0	0	330877	8.4583		Q	0
7	McCarthy, Mr. Timothy J	1	male	54	0	0	17463	51.8625	E46	S	0
8	Palsson, Master. Gosta Leonard	3	male	2	3	1	349909	21.075		S	0
9	Johnson, Mrs. Oscar	3	female	27	0	2	347742	11.1333		S	1
10	Nasser, Mrs. Nicholas	2	female	14	1	0	237736	30.0708		C	1

$$P\left(\frac{Survived}{Male \,\&\, Class \,\&\, Age \,\&\, Cabin \,\&\, Fare}\right)$$

Summary

- Understand conditional probability
- Introduced of Naïve bayes
- Advantages and disadvantages of Naïve bayes
- Hands on exercises

Himanshu Patel, Instructor
Saskatchewan Polytechnic
email: patelh@saskpolytech.ca
Mining building, saskatoon