CSE 5334: Dr. Chris Ding

Bayes Classification

From Bernoulli to Multinomial Distribution

Bayes Classification – Bayes Theorem

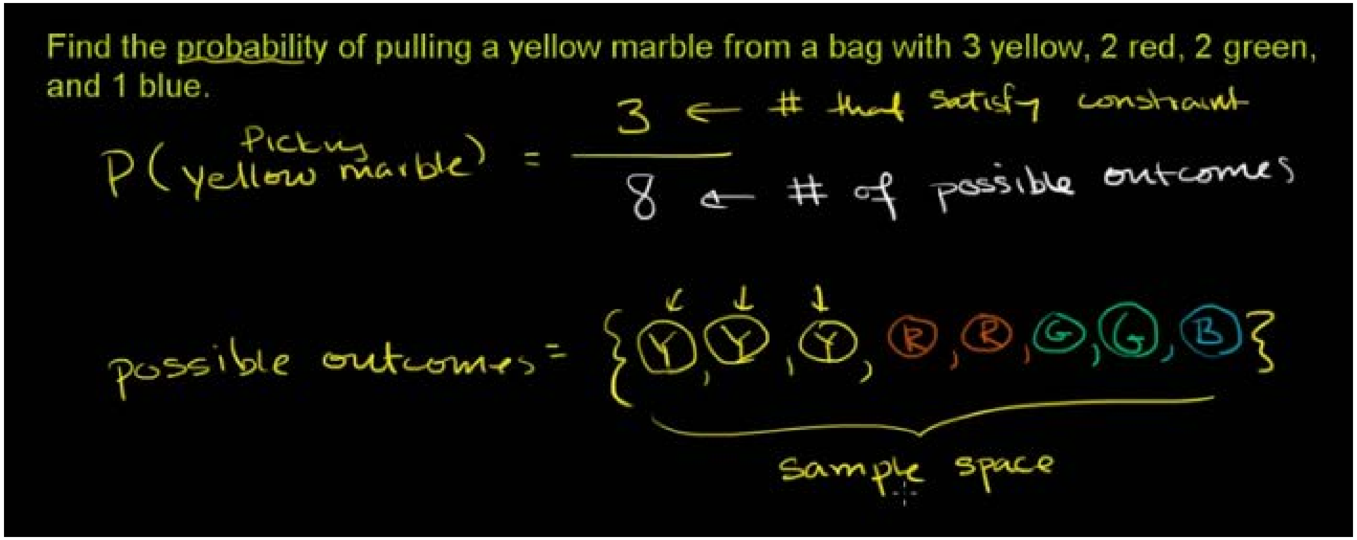
Naïve Bayes Classification – Attributes independence

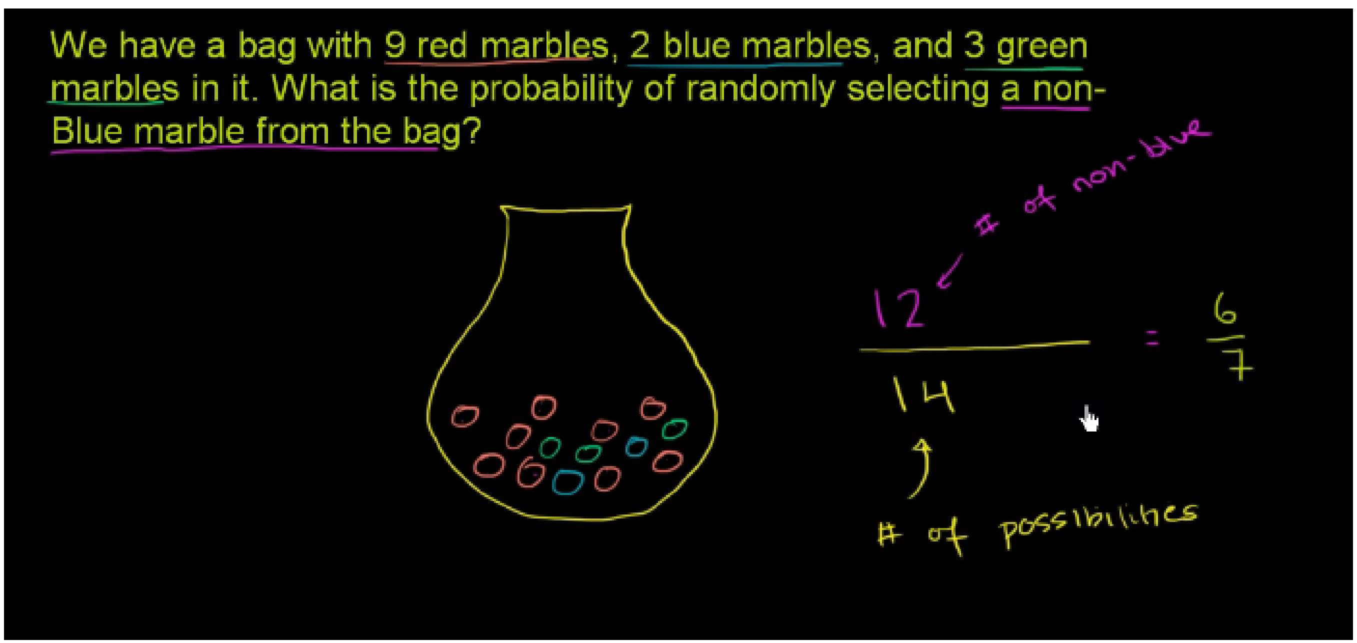
Reading: Textbook Sections 5.3, 5.3.1, 5.3.2, 5.3.3

From Bernoulli to Multinomial Distribution

- A brief review of probability, Bernoulli distribution, binomial distribution and multinomial distribution.
- Multinomial distribution plays vital important role in data mining/machine learning:
  - The basic model of English text, documents, fundamental theory for information retrieval, search engine, etc.
  - The basis for logistic regression and neural networks.
- An alternative (better) model of Naïve Bayes Classification

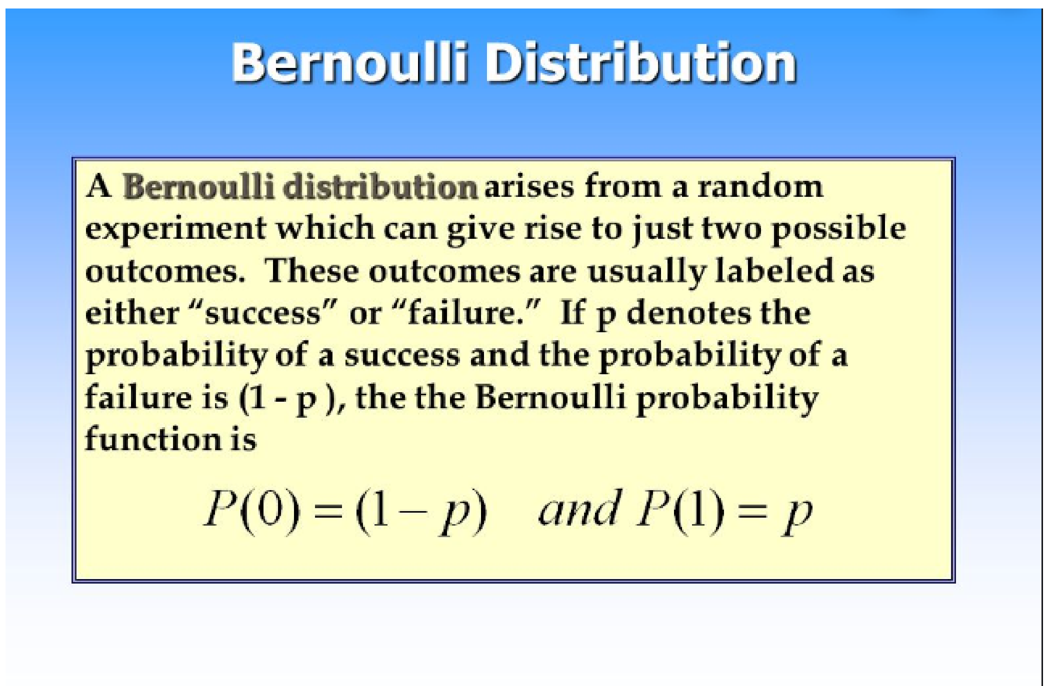
Probability = count possible outcomes satisfying requirements/constraints





Bernoulli distribution: Simplest probability distribution

Today is sunny or not-sunny.
Your team win or lose.
You throw a coin; it is head-up or head-down
Yow throw a die; the result is 6, or it is not 6 (which is 1 or 2 or 3 or 4 or 5)



Bernoulli Distribution

A **Bernoulli distribution** arises from a random experiment which can give rise to just two possible outcomes. These outcomes are usually labeled as either “success” or “failure.” If p denotes the probability of a success and the probability of a failure is (1 - p), the the Bernoulli probability function is

$$P(0) = (1 - p) \quad and\ P(1) = p$$

Binomial Distribution :

Y = X1 + … + Xn : sum of N independently identically distributed Bernoulli random variables

One experiment:

- the experiment consists of n independent trials, each with two mutually exclusive outcomes (**success** and **failure**)
- for each trial the probability of success is p (and so the probability of failure is $1 - p$)

Each such trial is called a **Bernoulli trial**.

Experiment: Throwing N identical coins, head-up/head-down

Experiment: Throwing one coin N times

Binomial Distribution Formula

$$P(x)=\binom{n}{x}p^{x}q^{n-x}=\frac{n!}{(n-x)!\,x!}p^{x}q^{n-x}$$

where

n = the number of trials (or the number being sampled)

x = the number of successes desired

p = probability of getting a success in one trial

$q = 1 - p$ = the probability of getting a failure in one trial

Example 1

Q. A coin is tossed 10 times. What is the probability of getting exactly 6 heads?

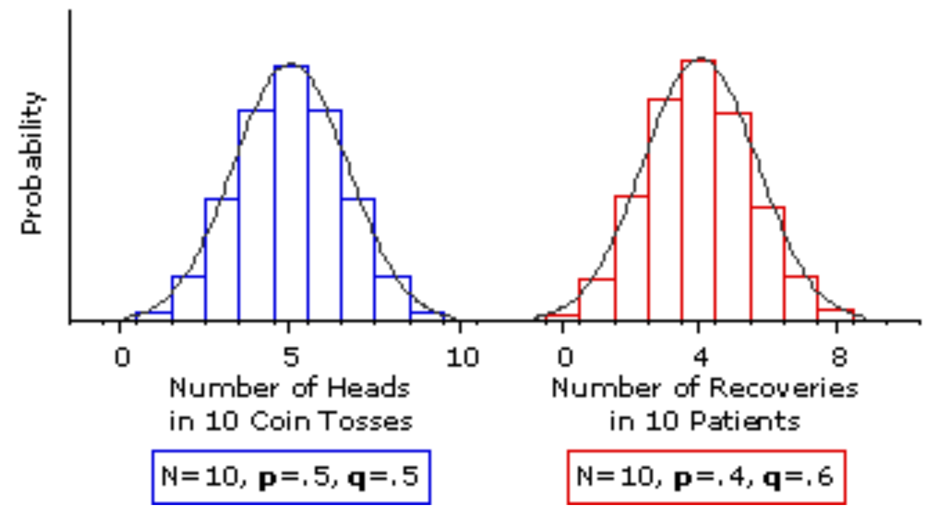
$$p = 0.5,\ q = 1 - p = 0.5,\ n = 10,\ x = 6$$

$$P(x=6) = \binom{10}{6} 0.5^6\ 0.5^{(10-6)} = 0.2051$$

$$P(x=5) = 0.2461$$

$$P(x=3) = P(x=7) = 0.1172$$

$$P(x=2) = P(x=8) = 0.0439$$



Example 3

60% of people who purchase sports cars are men. If 10 sports car owners are randomly selected, find the probability that exactly 7 are men.

$$p = 0.6,\ q = 1 - p = 0.4,\ n = 10,\ x = 7$$

$$P = \binom{10}{7} 0.6^7\ 0.4^{(10-7)} = 0.215$$

Multinomial Distribution

- The Binomial distribution can be extended to describe number of outcomes in a series of independent trials each having more than 2 possible outcomes.

- If a given trail can result in the k outcomes E_1, E_2, …, E_k with probabilities $p_1, p_2, \ldots, p_k$, then the probability distribution of the random variables $X_1, X_2, \ldots, X_k$, representing the number of occurrences for $E_1, E_2, \ldots, E_k$ in n independent trials is

$$p_{X_1,\ldots,X_k}(x_1,\ldots,x_k) = \frac{n!}{x_1!x_2!\cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$$

with $\sum_{i=1}^{k} x_i = n$, and $\sum_{i=1}^{k} p_i = 1.$

Example:

The distribution of blood types in the US is:

Type	O	A	B	AB
Probability	0.44	0.42	0.10	0.04

In a random sample of 10 Americans, what is the probability 6 have blood type O, 2 have type A, 1 has type B, and 1 has type AB?

$$P(X_1 = x_1, \ldots, X_k = x_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$$

$$P(X_1=6, X_2=2, X_3=1, X_4=1) = \frac{10!}{6!2!1!1!} 0.44^6 \, 0.42^2 \, 0.10^1 \, 0.04^1$$

$=0.01290$

Bayes Classification

Using Bayes Theorem (conditional probability) to obtain the class/label posterior probability of a data instance given its observed data (attributes/features)

Reading: Textbook Sections 5.3, 5.3.1, 5.3.2, 5.3.3

LIKELIHOOD
the probability of "B" being TRUE given that "A" is TRUE

PRIOR
the probability of "A" being TRUE

Data/Evidence

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

POSTERIOR
the probability of "A" being TRUE given that "B" is TRUE

The probability of "B" being TRUE

This formula is useful ONLY when A is class/hypothesis

Example: Test of Viral Infection

- A medical test for a viral infection. It is 95% reliable for infected patients and 99% reliable for the healthy ones：

- If a patient has the virus (event V), and the test shows that (event S) with probability $P\{S \mid V\} = 0.95$

- If a patient does not have the virus, the test confirms that with probability $P\{\overline{S} \mid \overline{V}\} = 0.99$

- A patient tests positive (the test shows that the patient has the virus).
- Does this means he has 95% probability of the virus?
- No!
- Because the question refers the probability that he has the virus and the test confirms that, i.e., P{V|S} . This quantity is not given directly in the statement of the problem.
- We compute P{V|S} using Bayes theorem.

Bayes' Rule

- Bayes Theorem (conditional probability):

$$P\{B \mid A\} = \frac{P\{A \mid B\}P\{B\}}{P\{A\}} = \frac{P\{A \mid B\}P\{B\}}{P\{A \mid B\}P\{B\} + P\{A \mid \overline{B}\}P\{\overline{B}\}}$$

Law of Total Probability

$$P\{A\} = \sum_{j=1}^{k} p\{\mathrm{A} \mid B_j\} P\{B_j\}$$

In case of two events (k=2),

$$P\{A\} = P\{A \mid B\} P\{B\} + P\{A \mid \overline{B}\} P\{\overline{B}\}$$

Medical Test Example cont.

- We need additional information: Suppose 4% of all the population are infected with the virus, P{V} = 0.04.
- Recall: $P\{S \mid V\} = 0.95$ $P\{\bar{S} \mid \bar{V}\} = 0.99$
- The desired (conditional) probability is

$$P\{V \mid S\} = \frac{P\{S \mid V\}P\{V\}}{P\{S \mid V\}P\{V\} + P\{S \mid \bar{V}\}P\{\bar{V}\}}$$

$$= \frac{(0.95)(0.04)}{(0.95)(0.04) + (1-0.99)(1-0.04)} = 0.7983$$

Test of Viral Infection - Conclusion

- Thus the probability of the patient has the virus is 79.83%, not 95%.
- Bayesian view:

This patient has 4% probability of been infected by the virus [because 4% of the population has the virus]. Because now he tested positive for the virus, his chance of virus increased to 79.83%.

This patient has 4% probability of been infected by the virus [because 4% of the population has the virus (prior probability)]. Because now he tested positive for the virus (new data evidence), his chance of virus increased to 79.83%.

Naïve Bayes Classification

Using Bayes Theorem (conditional probability) to obtain the class/label posterior probability of a data instance given its observed data (attributes/features)

5.3.3 Naïve Bayes Classifier

A naïve Bayes classifier estimates the class-conditional probability by assuming that the attributes are conditionally independent, given the class label y. The conditional independence assumption can be formally stated as follows:

$$P(\mathbf{X}|Y = y) = \prod_{i=1}^{d} P(X_i|Y = y), \tag{5.12}$$

where each attribute set $\mathbf{X} = \{X_1, X_2, \ldots, X_d\}$ consists of d attributes.

Probability of occurrence of X is equal to the product of the probability of occurrence of every attributes of X given the class of X

This says each class has a different multinomial distribution of attributes.

To classify a test record, the naïve Bayes classifier computes the posterior probability for each class Y:

$$P(Y|\mathbf{X}) = \frac{P(Y)\prod_{i=1}^{d} P(X_i|Y)}{P(\mathbf{X})}. \tag{5.15}$$

Since $P(\mathbf{X})$ is fixed for every Y, it is sufficient to choose the class that maximizes the numerator term, $P(Y)\prod_{i=1}^{d} P(X_i|Y)$.

the prior probability $P(Y)$

the class-conditional probabilities $\prod_i P(X_i|Y)$. = multinomial distribution of attributes for class Y

Compute probability of occurrence of each attributes for class Y="no"

Compute probability of occurrence of each attributes for class Y="yes"

Tid	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

(a)

P(Home Owner=Yes|No) = 3/7
P(Home Owner=No|No) = 4/7
P(Home Owner=Yes|Yes) = 0
P(Home Owner=No|Yes) = 1
P(Marital Status=Single|No) = 2/7
P(Marital Status=Divorced|No) = 1/7
P(Marital Status=Married|No) = 4/7
P(Marital Status=Single|Yes) = 2/3
P(Marital Status=Divorced|Yes) = 1/3
P(Marital Status=Married|Yes) = 0

For Annual Income:
If class=No: sample mean=110
sample variance=2975
If class=Yes: sample mean=90
sample variance=25

(b)

Figure 5.10. The naïve Bayes classifier for the loan classification problem.

Standard multinomial distribution parameter estimation:

$$P(x_i|Y = y)^{\text{MLE}} = p_{i,y}^{\text{MLE}} = \frac{n_{i,y}}{N_y}$$

where $n_{i,y}$ is the number of training examples in class y where attribute x_i occurs, N_y is the number of training examples in class y.

Laplace smoothed multinomial distribution parameter estimation:

See 2nd Edition Textbook p.224

$$P(x_i|Y = y)^{\text{smoothed}} = p_{i,y}^{\text{smoothed}} = \frac{n_{i,y} + 1}{N_y + \nu_y}$$

where ν is the total number of attributes in class y.

In most applications, we use Laplace smoothed parameter estimation