2.4.4. Complex set

In high school, a complex number is defined by giving its form ($z = a + bi$; $a, b \in \mathbb{R}$) with the appearance of an imaginary number i that satisfies $i^2 = -1$.

In algebraic structure, the complex set $\mathbb{C}$ is constructed from the set $\mathbb{R}^2 = \{(a,b) | a, b \in \mathbb{R}\}$ with the addition and multiplication operations defined: For all $(a,b), (c,d) \in \mathbb{R}^2$ we have

$$(a,b) + (c,d) = (a+c, b+d);\ (a,b).(c,d) = (ac - bd; ad + bc).$$

The set $\mathbb{R}^2$ together with the two operations is a field, which is called a *complex field* $\mathbb{C}$.

It is clear that the map $f: \mathbb{R} \to \mathbb{C}, a \mapsto (a, 0)$.

is a field monomorphism. Hence, we can identical element $a \in \mathbb{R}$ with element $(a,0) \in \mathbb{C}$. This leads to $\mathbb{R} \subset \mathbb{C}$.

Let $i = (0,1) \in \mathbb{C}$. We have $i^2 = (0,1).(0,1) = (-1,0) = -1$ and every $x = (a,b) \in \mathbb{C}$ can be written $x = (a,b) = (a,0) + (0,b) = (a,0) + (b,0).(0,1) = a + b.i$.

From the perspective of field extension theory, $\mathbb{C}$ is a simple extension of $\mathbb{R}$ with algebraic element i on $\mathbb{R}$, that is, $\mathbb{C} = \mathbb{R}(i)$.

3. CONCLUSION

In this article, I present an explanation for solutions to product equation with a variable, first-degree equation with an unknown, problems of finding differences between two numbers or construction of number sets from the perspective of algebraic structures.

REFERENCES

1. M. F. Atiyah, I. G. Macdonald (1969), *Introduction to Commutative Algebra*, Massachusetts : Addison-Wesley Publishing.
2. Patrict Morandi (1996), *Field and Galois Theory*, Springer.
3. Bui Huy Hien (2005), *General Algebra Exercises*, Education Publishing.
4. Hoang Xuan Sinh (2015), *General Algebra*, Education Publishing.
5. Hoang Quoc Viet, Dam Van Nhi (2017), *Number Theory and Polynomial Bases,* University of Education Publishing.

XEM XÉT MỘT SỐ BÀI TOÁN SƠ CẤP DƯỚI GÓC NHÌN CỦA CÁC CẤU TRÚC ĐẠI SỐ

Tóm tắt: *Bài viết này xem xét bản chất của một số bài toán sơ cấp dưới góc nhìn của cấu trúc đại số nhóm, vành, trường, từ đó giải thích cơ sở lí luận cho lời giải của bài toán đó.*

Từ khóa*: Cấu trúc đại số, bài toán sơ cấp, mở rộng trường.*

SOME PROBABILITY MODELS FOR ARTIFICIAL INTELLIGENCE

Nguyen Van Tuan

Hanoi Metropolitan University

Abstract: *In this article, we will show some probability models for artificial intelligence. They are Bayesian and linear regression models. Each model is illustrated by its own example.*

Keywords: *Probability model; AI; Bayesian rule; Bayesian networks; linear regression model.*

Received: 18 March 2020

Accepted for publication: 20 April 2020

Email: tuannv@daihocthudo.edu.vn

1. INTRODUCTION

In recent years mathematicians have been interested in studying the application of probability models to Artificial Intelligence (AI). This is because of the probability of having great applications in AI and life. The main researching areas are generalizations of using probabilities in AI or using probabilities for specific issues of AI. For example, M. Kukacka [2] gave an overview of Bayesian Methods in Artificial Intelligence; Sunghae [6] has studied a probability learning model for constructing artificial minds; Credit risk analysis using machine and deep learning models was introduced by Peter and others [4],…

In this paper, we have applied the probability for AI. This work is built as follow: in Section 2, Bayesian models for AI and their application are presented. The Linear regression models in prediction are established in Section 3. In the last Section, Section 4, conclusion is presented.

2. CONTENT

2.1. Bayesian models for AI

2.1.1. Bayesian rule and applications

1. Bayesian rule

Assume that events $\{H_i\}_{i=1}^{n}$ form a partition of the sample space Ω, i.e.

$$\Omega = \bigcup_{i=1}^{n} H_i,\ \ H_i \cap H_j = \emptyset, i \neq j.$$

Then an event A in Ω, we have

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{\sum_{i=1}^{n} P(H_i)P(A|H_i)}, i = 1, 2, \dots, n.$$

Applying Bayesian rule, we can solve many problems of AI that have applications in real-life.

2. Some examples

Example 1. In a factory machines (I), (II) and (III) are all producing springs of the same length. Of their production, machines (I), (II) and (III) produce 2%, 1% and 3% defective springs, respectively. Of the total production of springs in the factory, machine (I) producces 35%, machine (II) producces 25%, machine (III) producces 40%.

If one of the factory's springs is randomly selected, the possibility of obtaining a spring is made by which machine.

Solution

Call A, B, C as random events to choose the spring made by machines (I), (II) and (III) produce. Then {A, B, C} form a complete system.

Let H be an incident that selects a factory product.

We have $P(A) = 0.35$; $P(A) = 0.35$; $P(A) = 0.35$; $P(H|A) = 0.02$; $P(H|A) = 0.02$; $P(H|A) = 0.02$.

If one spring is selected at random from the total springs, the propability that it is defected equals:

$$P(H) = P(A)P(H|A) + P(B)P(H|B) + P(C)P(H|C) = 0.0218.$$

Applying Baysian rule, we get

$$P(A|H) = \frac{P(A)P(H|A)}{P(H)} = \frac{70}{215};$$

$$P(B|H) = \frac{P(B)P(H|B)}{P(H)} = \frac{25}{215};$$

$$P(C|H) = \frac{P(C)P(H|C)}{P(H)} = \frac{120}{215}.$$

So most likely the spring obtained by the machine (III).

Example 2. The probability of a certain medical test being positive is 90%, if a patient has flu. 1% of the population has the disease, and the test records a false positive 5% of the time. If a man in that region received a positive test, how much is his probability of having flu?

Solution.

Let H be the event of selecting a person who has the flu in that area.

Let A be the positive test event for a person with the flu in that area.

By hypothesis, we have

$$P(H) = 0.01; P(A|H) = 0.90;\ P(A|\overline{H}) = 0.05.$$

Because events $\{H_i\}_{i=1}^{n}$ form a partition of the space Ω, so we get:

$$P(A) = P(H)P(A|H) + P(\overline{H})P(A|\overline{H}) = \ 0.0585.$$

Applying Bayesian rule, we have

$$P(H|A) = \frac{P(H)P(A|H)}{P(A)} = \frac{0.01 \times 0.9}{0.0585} = 0.15$$

So if a man in that region received a positive test, then his probability of having flu is 0.15.

2.2. Bayesian netwwork

Bayes networks have many applications in AI. Here, we present some basis for Bayes networks application in AI.

1. Definition. A Bayesian network (BN) is defined by the following elements [2]:

• A set of nodes, where each node represents a single variable;

• A set of directed connections on these nodes, forming a directed acyclic graph, where a link specifies a dependence relationship between variables;

• A conditional probability table (CPT) for each node in the graph, specifying a probability distribution of the corresponding variable conditioned by its parents in the graph (i.e. $P(X_i|Parents\ (X_i))$).

We can retrieve the probability of any event in a system described by a Bayesian network using the following formula:

$$P(x_1, \dots, x_n) = \prod_{i=1}^{n} (x_i|\text{Parents}(X_i)),$$

where parents(X_i) denotes the specific values of the X_i 's parent variables. This implies that the Bayesian network fully describes the full joint distribution of the system.

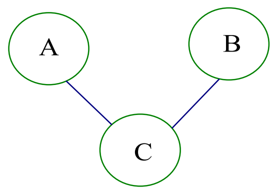
2. Some rules of Bayesian networks

Let A, B, C are random variables. We say that A and B independent with given C, if we know C, evidence of B does not change the likelihood of A.

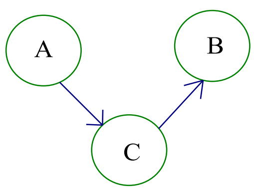
If A and B independent given C then $P(A|B,C) = P(A|C)$. When we have

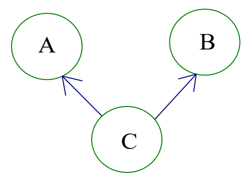
i) $P(B|A,C) = P(B|C)$;
ii) $P(A,B,C) = P(A|C)P(B|C)P(C)$;
iii) Graphs of A, B and C as following:

+ Undirected graphs



+ Directed acyclic graphs



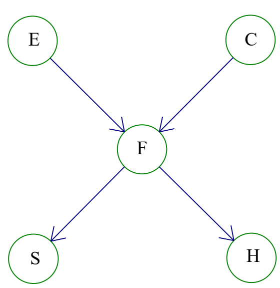


3. Example of causal models

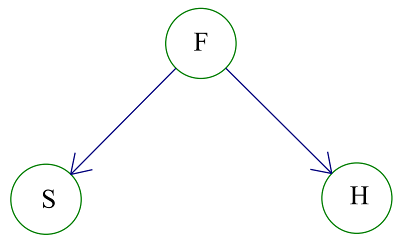
We can use graphical to represent causal models. Let E, C, F, S, H are random variables, where E: electric; C: cigaratte; F: fire; S: smoke; H: heat.

We know that, fire can be caused by an electrical problem or by a cigarette. Smoke and Heat are results of firing.

Smoke and Heat is the consequence of burning. Therefore, we have the following causal model:



Now we consider a problem: Let we have a graphical model:



Each factor is described by a conditional probability table

F	$P(S)$	$P(\bar{S})$
F	0.01	0.99
T	0.9	0.1

F	$P(H)$	$P(\bar{H})$
F	0.001	0.999
T	0.99	0.01

$P(F)$	$P(\bar{F})$
0.1	0.9

Calculate the probability of fire when smoke has occurred.

Solution.

From the hypothesis, we have

$$\begin{cases} P(F) = 0.1 \\ P(\bar{F}) = 0.9, \end{cases} \quad \begin{cases} P(S|\bar{F}) = 0.01 \\ P(\bar{S}|\bar{F}) = 0.99 \\ P(S|F) = 0.9 \\ P(\bar{S}|F) = 0.1, \end{cases} \quad \begin{cases} P(H|\bar{F}) = 0.001 \\ P(\bar{H}|\bar{F}) = 0.999 \\ P(H|F) = 0.99 \\ P(\bar{H}|F) = 0.01. \end{cases}$$

The probability of fire and smoke is

$$P(S,H) = \sum_H P(F,S,H) = \sum_H P(H|F)P(S|F)P(F)$$

$$= P(H|F)P(S|F)P(F) + P(\bar{H}|F)P(S|F)P(F)$$

$$= 0.9 \times 0.99 \times 0.1 + 0.90 \times 0.01 \times 0.1$$

$$= 0.09.$$

The probability of seeing smoke is

$$P(S) = \sum_F \sum_H P(F,S,H) = \sum_F \sum_H P\,(H|F)P(S|F)P(F)$$

$$= P(H|F)P(S|F)P(F) + P(\bar{H}|F)P(S|F)P(F) +$$

$$+ \; P(H|\bar{F})P(S|\bar{F})P(\bar{F}) + P(\bar{H}|\bar{F})P(S|\bar{F})P(\bar{F}) = 0.99 \times 0.9 \times 0.1 +$$
$$0.01 \times 0.9 \times 0.1 + 0.001 \times 0.01 \times 0.9 + 0.999 \times 0.01 \times 0.9 = 0.099.$$

So the probability of fire if we see smoke it is defined

$$P(F|S) = \frac{P(F,S)}{P(S)} = \frac{0.09}{0.099} = 0.92$$

2.3. Some linear regression models for AI

Linear regression can be applied in AI [6]. In this section we will build linear regression models and get some illustrative examples.

2.3.1. Linear regression models for AI

We observe paired data points $\{(x_i, y_i)\}_{i=1}^{n}$, where assume that as a function of x_i, each y_i is generated by using some true underlying line $Y = \beta_0 + \beta_1 X$ that is evalute at x_i. Formally,

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1, \dots, n. \quad (3.1)$$

We will model ε_i as being Gaussian: $\varepsilon \sim N(0, \sigma^2)$. We find β_0, β_1 by solving the following optimization problem:

$$min_{\beta_0,\beta_1} \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2. \quad (3.2)$$

Given a set of points $\{(x_i, y_i)\}_{i=1}^{n}$, the solution is [6]:

$$\widehat{\beta_1} = \frac{\overline{XY} - \overline{X}\,\overline{Y}}{\overline{X}^2 - (\overline{X})^2} = r_{XY}\frac{\sqrt{MS_Y}}{\sqrt{MS_X}},$$

$$\widehat{\beta_0} = \overline{Y} - \widehat{\beta_1}\overline{X},$$

where

$$MS_X = \overline{X}^2 - \left(\overline{X}\right)^2, MS_Y = \overline{Y}^2 - \left(\overline{Y}\right)^2, \overline{XY} = \frac{\sum_{i=1}^{n} x_i y_i}{n}, \overline{X}^2 = \frac{\sum_{i=1}^{n} {x_i}^2}{n}, \overline{Y}^2 = \frac{\sum_{i=1}^{n} {y_i}^2}{n}.$$

Example 1. On 100 field plots of the same size, people apply different amounts of fertilizer. Then, study the relationship between fertilizer and yield. The result is given in the table below (X is the amount of fertilizer, Y is the yield).

Y \ X	1	2	3	4	5
14	10	8			
15		12	7		
16			28	6	
17				8	9
18					12

a) If the amount of manure is 3.44, what is the estimate yield?

b) Estimate yield if the amount of fertilizer is 5.65.

Solution. We have

$$\sum_{i=1}^{100} x_i = 1 \times 10 + (8+12) \times 2 + (7+28) \times 3 + (6+8) \times 4 + (9+12) \times 5 = 316.$$

$$\overline{X} = \frac{\sum_{i=1}^{100} x_i}{n} = \frac{316}{100} = 3.16.$$

$$\sum_{i=1}^{100} y_i = (1+8) \times 14 + (12+7) \times 15 + (6+28) \times 16 + (8+9) \times 17 + 12 \times 8$$
$$= 1586.$$

$$\overline{y} = \frac{\sum_{i=1}^{100} y_i}{n} = \frac{1586}{100} = 15.86.$$

$$\sum_{i=1}^{100} {x_i}^2 = 1154; \sum_{i=1}^{100} {y_i}^2 = 25308; \sum_{i=1}^{100} x_i y_i = 5156;$$

$$100 \sum_{i=1}^{100} x_i y_i - \sum_{i=1}^{100} x_i \left(\sum_{i=1}^{100} y_i \right) = 14424;$$

$$\sqrt{MS_X} = \sqrt{100 \sum_{i=1}^{100} {x_i}^2 - \left(\sum_{i=1}^{100} x_i\right)^2} = 124.67;$$

$$\sqrt{MS_Y} = \sqrt{100 \sum_{i=1}^{100} {y_i}^2 - \left(\sum_{i=1}^{100} y_i\right)^2} = 124.11;$$

$$S\,Ir_{XY} = \frac{14424}{124.67 \times 124.11} = 0.93.$$

So we get

$$\hat{Y} = \widehat{\beta_0} + \widehat{\beta_1} X \Rightarrow \hat{Y} = 12.93 + 0.93X.$$

a) If $X = 3.44$ then $\hat{Y} = 16.13$;

b) If $X = 5.65$ then $\hat{Y} = 18.19$.

Example 2. The same pill was given to 5 patients of different ages. Study the time to completely disintegrate the drug in each person's body. Specific results are as follows:

X: age (year)	Y: Decomposition time (minute)
30	15
25	28
65	30

50	22
40	24

a) What conclusions can be drawn about the relationship between the time of drug decomposition and ages.

b) What is the time predict for drug decomposition by 42 years old?

Solution.

We have

$$\sum_{i=1}^{5} x_i = 210; \sum_{i=1}^{5} x_i = 119; \sum_{i=1}^{5} x_i\, y_i = 5160; \sum_{i=1}^{5} {x_i}^2 = 9850.$$

$$\widehat{\beta_1} = \frac{n\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} {x_i}^2 - (\sum_{i=1}^{n} x_i)^2} = \frac{5 \times 5160\ -\ 210 \times 119}{5.9850\ -\ 210^2} = 0.16;$$

$$\widehat{\beta_0} = \frac{\sum_{i=1}^{n} y_i - \widehat{\beta_1}\sum_{i=1}^{n} x_i}{n} = \frac{5 \times 5160\ -\ 210 \times 119}{5.9850\ -\ 210^2} = 17.08.$$

So we get

$$\hat{Y} = 17.08 + 0.16X. \tag{3.3}$$

By (3.3) we say that

a) Every 10 years of age, the decomposition time will increase by 1.6 minutes.

b) The time for decomposition of 42-year-old patient is 23.8 minutes.

2.3.2. Multiple linear regressions

Let $Y, X_2, \dots, X_p$ are p radom variables, where $p\ \in \boldsymbol{N}^*,\ p > 2.$ We find linear model of Y for $X_2, \dots, X_p$ as following:

$$f_Y(X_2, \dots, X_p) =\ \beta_1 +\ \beta_2 X_2 + \dots +\ \beta_p X_p,$$

which observation matrix

$$\begin{pmatrix} Y_1 & X_{12} & X_{13} \dots & X_{1p} \\ Y_2 & X_{22} & X_{23} \dots & X_{2p} \\ \dots & \dots & \dots\ \dots & \dots \\ Y_n & X_{n2} & X_{n3} \dots & X_{np} \end{pmatrix}.$$

We get regression model

$$Y_i =\ \beta_1 +\ \beta_2 X_{i2} + \dots +\ \beta_p X_{ip} +\ \varepsilon_i, \tag{3.4}$$

where $\varepsilon \sim N(0, \sigma^2)$.

Let

$$Y = [Y_1, \dots, Y_n]^T;\ X = \begin{pmatrix} 1 & X_{12} & X_{13} \dots & X_{1p} \\ 1 & X_{22} & X_{23} \dots & X_{2p} \\ \dots & \dots & \dots \dots & \dots \\ 1 & X_{n2} & X_{n3} \dots & X_{np} \end{pmatrix};\ X_j = [X_{1j}, \dots, X_{nj}]^T, j = 2, \dots, p; \quad (3.5)$$

$$\beta = [\beta_1, \dots, \beta_p]^T;\ \varepsilon = [\varepsilon_1, \dots, \varepsilon_n]^T.$$

$$Y = [Y_1, \dots, Y_n]^T;\ X = \begin{pmatrix} 1 & X_{12} & X_{13} \dots & X_{1p} \\ 1 & X_{22} & X_{23} \dots & X_{2p} \\ \dots & \dots & \dots \dots & \dots \\ 1 & X_{n2} & X_{n3} \dots & X_{np} \end{pmatrix};\ X_j = [X_{1j}, \dots, X_{nj}]^T, j = 2, \dots, p;$$

Using notations (3.5) and (3.6), we can write model (3.4) as following:

$$Y = X\beta + \varepsilon.$$

To determine β we have to solve the optimization problem

$$min_\beta \sum_{i=1}^{n} (y_i - X_i\beta)^2 \quad (3.7)$$

We can use some basic linear algebra to solve this problem and find the result:

$$\hat{\beta} = (X^T X)^{-1} X^T Y. \quad (3.8)$$

Example 3. A Corporation has 21 stores in provinces and cities. Let Y be the revenue of the Corporations (10^3USD), X_1 is the population (thousand), X_2 is the average income per person (10^3 USD) in the provinces and cities where the business is located. Assume the following table of business data is available:

Provinces and cities	X_1 (Thousand)	X_2 (10^3USD)	Y (10^3USD)
1.	68.5	16.7	174.4
2.	45.2	16.8	164.4
3.	91.3	18.2	224.2
4.	47.8	16.3	154.6
5.	46.9	17.3	181.6
6.	66.1	10.2	207.5
7.	49.5	15.9	152.8
8.	52.0	17.2	163.2
9.	48.9	16.6	145.4
10.	38.4	16.0	137.2
11.	87.9	18.3	241.9
12.	72.8	17.1	191.1
13.	88.4	17.4	232.0

14.	42.9	15.8	145.3
15.	52.5	17.8	161.1
16.	85.7	18.4	209.7
17.	43.1	16.5	146.4
18.	51.7	16.3	144.0
19.	89.6	18.1	232.6
20.	82.7	19.1	224.1
21.	52.3	16.0	166.5

Predict the revenue of this Corporation.

Solution.

Use model (3.4) which $p = 3$ to solve this problem.

We have

$$X = \begin{pmatrix} 1 & 68.5 & 16.7 \\ 1 & 42.2 & 16.8 \\ \dots & \dots & \dots \\ 1 & 52.3 & 16.0 \end{pmatrix}; Y = \begin{pmatrix} 174.4 \\ 164.4 \\ \dots \\ 166.5 \end{pmatrix}.$$

After calculating we obtained

$$X^T X = \begin{pmatrix} 21.0 & 1302.4 & 360.0 \\ 1302.4 & 87707.9 & 22609.2 \\ 360 & 22609.2 & 6190.3 \end{pmatrix}, X^T Y = \begin{pmatrix} 3820 \\ 249643 \\ 66073 \end{pmatrix},$$

$$(X^T X)^{-1} = \begin{pmatrix} 29.5740 & 0.0718 & -1.9820 \\ 0.0718 & 0.00037 & -0.0055 \\ -1.9820 & -0.0055 & 0.1356 \end{pmatrix}$$

And

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \begin{pmatrix} -68.609 \\ 1.455 \\ 9.488 \end{pmatrix}.$$

So we have the result

$$\hat{Y} = -68.609 + 1.455 X_1 + 9.488 X_2. \tag{3.9}$$

From (3.9) we can predict the results:

a) If the average income is constant and the population increases by one thousand people, the sale turnover of the Corporation increases by 1455 USD;

b) If the population is constant and the income per capita increases by one thousand dollars, the sale turnover increases by 9488 USD.

3. CONCLUSION