Estimating Conditional Probabilities for Continuous Attributes

Estimating Conditional Probabilities for Continuous Attributes

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp^{-\frac{(x_i - u_{ij})^2}{2\sigma_{ij}^2}}$$

For example, $P(Income = 120|No) = ?$

$$\mu = \frac{125 + 100 + 70 + 120 + 60 + 220 + 75}{7} = 110$$

7 → มีทั้งหมดกี่ตัว

$$\sigma^2 = \frac{(125-110)^2 + (100-110)^2 + \ldots + (75-110)^2}{6} = 2975$$

→ เอาเลข $(income|NO - \mu)^2$

$\sigma = \sqrt{2975} \simeq 54.54$ → σ เอาเข้า $\sqrt{}$

$$P(Income = 120|No) = \frac{1}{\sqrt{2\pi}(54.54)} \exp^{-\frac{(120-110)^2}{2\times 2975}} \simeq 0.0072$$

สูตร $P(X_i \overset{120}{=} x_i | Y \overset{NO}{=} y_j) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp^{-\frac{(x_i - u_{ij})^2}{2\sigma_{ij}^2}}$ (54.54)

	binary	categorical	continuous	class
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

Exercises

Class-Labeled Training Tuples from the *AllElectronics* Customer Database

RID	age	income	student	credit_rating	Class: buys_computer
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

Predict class label of

$\mathbf{X} = (age = youth, income = medium, student = yes, credit\_rating = fair)$

Predict class label of

$\mathbf{X} = (age = youth, income = medium, student = yes, credit\_rating = fair)$

สูตร $P(Y|X) = \frac{(P(X|Y))P(Y))}{\boxed{P(X)}}$ → เราสามารถ ignore ได้

	$P(Y=yes\|x)$ vs	$P(Y=NO\|X)$
	$P(Y=yes) = 9/14$	$P(Y=NO) = 5/14$
age = youth	2/9	3/5
income = medium	4/9	2/5
student = yes	6/9	1/5
credit_rating = fair	6/9	2/5

สูตร $P(X|Y=yes)P(Y=yes)$
$= (\frac{2}{9})(\frac{4}{9})(\frac{6}{9})(\frac{6}{9})(\frac{9}{14})$
$= 0.0282$

$P(X|Y=NO)P(Y=NO)$
$(\frac{3}{5})(\frac{2}{5})(\frac{1}{5})(\frac{2}{5})(\frac{5}{14})$
$= \frac{6}{875} = 0.0069$

The class of X = (age=youth, income=medium, student=yes, credit_rating=fair) as "Yes"

Predit X = (age=senior, income=medium, student=no, credit_rating= fair)

	$P(Y=Yes) = 9/14$	$P(Y=NO) = 5/14$
age=senior	3/9	2/5
income=medium	4/9	2/5
student=no	3/9	4/5
credit_rating=fair	6/9	2/5

$P(X|Y=Yes)P(Y=Yes)$
$(\frac{3}{9})(\frac{4}{9})(\frac{3}{9})(\frac{6}{9})(\frac{9}{14})$
$= 0.021$

$P(X|Y=NO)(Y=NO)$
$(\frac{2}{5})(\frac{2}{5})(\frac{4}{5})(\frac{2}{5})(\frac{5}{14})$
$= 0.018$

$\therefore$ $P(Y=Yes) > P(Y=NO)$, the record is classfield as "Yes"

Linear Regression(สมการเส้นตรง)

Simple Linear Regression

equation:

→ สมการเส้นตรง

$$\boxed{y = \beta_0 + \beta_1 x}$$

where:

y is the dependent variable
x is the independent variable
β_0 is the y-intercept → จุดตัดแกน y, ค่าคงที่
β_1 is the slope ค่าสัมประสิทธิ์ของตัวแปร x
ค. ชันของเส้นตรง

Multiple Linear Regression

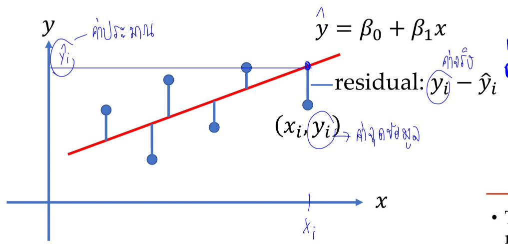
equation:

จุดตัดแกน y, ค่าคงที่

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

where y is the dependent variable
$x_1, x_2, \ldots, x_n$ are the independent variables
$\beta_1, \beta_2, \ldots, \beta_n$ are the regression coefficients สัมประสิทธิ์
β_0 is the y-intercept (constant term)

Ordinary Least Squares (OLS)



B0 = จุดตัดแกน y, ค่าคงที่
B1 = ค่าสัมประสิทธิ์, ค.ชัน
yi = ค่าจริง
$\hat{y}$ = ค่าประมาณ

- The objective of OLS is to find β_0 and β_1 that minimize sum of squared residuals; that is

$$\text{Minimize} \sum_i \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2$$

SSR : Sum of squared residuals = ผลรวมของ error

$$SSR = \sum_i (y_i - \hat{y}_i)^2 = \sum_i \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2$$

- The solutions of above minimization are: n = จำนวนข้อมูล

$$\beta_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

↳ ค่าเฉลี่ย y (ผลรวมของ y / จำนวนข้อมูล)

n = จำนวนข้อมูล

$\bar{y}$ = $\dfrac{\text{ผลรวมของ y}}{\text{จำนวนข้อมูล}}$

- where $\bar{x}$ and $\bar{y}$ are means of $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ respectively.

Example :

- Find a linear regression equation for the following data.

x	y
2	3
4	7
6	5
8	10

1. หาสมการ SSR
2. mean squared error
3. √ mse

① หา x^2, xy, ผลรวม

- Solution: To find the linear regression equation we need to find the value of Σx, Σy, Σx², and Σxy.

4 record

	x	y	x²	xy
	2	3	4	6
	4	7	16	28
	6	5	36	30
	8	10	64	80
ผลรวม	Σx = 20	Σy = 25	Σx² = 120	Σxy = 144

② เข้าสูตรตามนี้

- We get

Σ ผลรวม

$$\hat{\beta}_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} = \frac{4 \times 144 - 20 \times 25}{4 \times 120 - 400} = 0.95$$

↳ จำนวนข้อมูล, จำนวน record

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{25}{4} - 0.95 \times \frac{20}{4} = 1.5$$

- Thus, the equation of linear regression is y = 1.5 + 0.95x

y = B1 + B0 X

Evaluation Metrics for Regression Models

① Mean Squared Error (MSE)

SSR

$$MSE = \frac{1}{n}\boxed{\sum_i (y_i - \hat{y}_i)^2}$$

ค่าประมาณ

Example:

Example

มาจาก y = 1.5 + 0.95x

$\hat{y} = 1.5 + 0.95x$
$= 1.5 + 0.95(2)$

4 record

x	y	
2	3	
4	7	
6	5	
8	10	
ผลรวม $\sum x = 20$	$\sum y = 25$	$\sum$

y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
3	3.4	-0.4	0.16
7	5.3	1.7	2.89
5	7.2	-2.2	4.84
10	9.1	0.9	0.81

$$MSE = \frac{1}{n}\sum_i (y_i - \hat{y}_i)^2 = \frac{1}{4}(0.16 + 2.89 + 4.84 + 0.81) = 2.175$$

Note: $\hat{y}$ มาจากสมการเส้นตรง

② Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_i (y_i - \hat{y}_i)^2}$$

Example:

Example

y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
3	3.4	-0.4	0.16
7	5.3	1.7	2.89
5	7.2	-2.2	4.84
10	9.1	0.9	0.81

$$RMSE = \sqrt{\frac{1}{n}\sum_i (y_i - \hat{y}_i)^2} = \sqrt{\frac{1}{4}(0.16 + 2.89 + 4.84 + 0.81)} = \sqrt{2.175} = 1.47$$

③ Coefficient of Determination

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

SSR

Example.

Example

$\bar{y} = \frac{3 + 7 + 5 + 10}{4}$
$= \frac{25}{4} = 6.25$

y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$	$y - \bar{y}$	$(y - \bar{y})^2$
3	3.4	-0.4	0.16	-3.25	10.5625
7	5.3	1.7	2.89	0.75	0.5625
5	7.2	-2.2	4.84	-1.25	1.5625
10	9.1	0.9	0.81	3.75	14.0625

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{(0.16 + 2.89 + 4.84 + 0.81)}{(10.5625 + 0.5625 + 1.5625 + 14.0625)} = 0.67$$

Note: $\bar{y}$ มาจาก ผลรวมของค่า y / จำนวน record y

Limits and Derivatives

behavior of the function = การเปลี่ยนแปลงของค่า function

Properties of Limits

- **Theorem** (Some Basic Limits): Let a and b are real numbers and let n be a positive integer.

1. $\lim_{x \to a} b = b$
2. $\lim_{x \to a} x = a$
3. $\lim_{x \to a} x^n = a^n$

Properties of Limits

- **Theorem** (Limit Laws): Suppose that c is a constant and the limits $\lim_{x \to a} f(x)$ and $\lim_{x \to a} g(x)$ exist. Then

1. $\lim_{x \to a}[f(x) + g(x)] = \lim_{x \to a} f(x) + \lim_{x \to a} g(x)$
2. $\lim_{x \to a}[f(x) - g(x)] = \lim_{x \to a} f(x) - \lim_{x \to a} g(x)$
3. $\lim_{x \to a}[cf(x)] = c \lim_{x \to a} f(x)$
4. $\lim_{x \to a}[f(x) \times g(x)] = \lim_{x \to a} f(x) \times \lim_{x \to a} g(x)$
5. $\lim_{x \to a} \dfrac{f(x)}{g(x)} = \dfrac{\lim_{x \to a} f(x)}{\lim_{x \to a} g(x)}$ if $\lim_{x \to a} g(x) \neq 0$

Example :

- Evaluate the following limits

ⓐ $\lim_{x \to 5} (2x^2 - 3x + 4)$

$= \lim_{x \to 5} 2x^2 - \lim_{x \to 5} 3x + \lim_{x \to 5} 4$

$= 2\lim_{x \to 5} x^2 - 3\lim_{x \to 5} x + \lim_{x \to 5} 4$

$= 2(5)^2 - 3(5) + 4$

$= 50 - 15 + 4$

$= 39$

ⓑ $\lim_{n \to -2} \dfrac{x^3 + 2x^2 - 1}{5 - 3x}$

$= \dfrac{\lim_{n \to -2} x^3 + \lim_{n \to -2} 2x^2 - \lim_{n \to -2} 1}{\lim_{n \to -2} 5 - \lim_{n \to -2} 3x}$

$= \dfrac{\lim_{n \to -2} x^3 + 2\lim_{n \to -2} x^2 - \lim_{n \to -2} 1}{\lim_{n \to -2} 5 - 3\lim_{n \to -2} x}$

$= \dfrac{(-2)^3 + 2(-2)^2 - 1}{5 - 3(-2)}$

$= \dfrac{(-8) + 2(4) - 1}{5 - (-6)} =$

$= -\dfrac{1}{11}$

Coutinuity

- **Theorem:** $\lim_{x \to a} f(x) = L$ if and only if $\lim_{x \to a^-} f(x) = L$ and $\lim_{x \to a^+} f(x) = L$

- **Definition:** A function f is **continuous** at a number a if $\lim_{x \to a} f(x) = f(a)$

- Notice that it implicitly requires three things if f is continuous at a:

1. $f(a)$ is defined
2. $\lim_{x \to a} f(x)$ exists
3. $\lim_{x \to a} f(x) = f(a)$

Rates of Change

The (instantaneous) คือ rate of change of y with respect to x

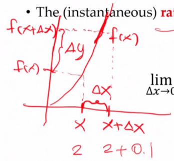
Δx แทนเป็น h (อนุพันธ์)

อนุพันธ์ = อัตราการเปลี่ยนแปลงของ y เทียบกับ x

- The (instantaneous) **rate of change of y with respect to x** is

- The (instantaneous) **rate of change of y with respect to x** is

$\Delta x = h$
↓
อนุพันธ์



$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \boxed{\lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}}$$

$$\boxed{\frac{\Delta y}{\Delta x} = \frac{f(x + \Delta x) - f(x)}{\Delta x}}$$

Derivatives

$$\boxed{f'(x)} = \frac{dy}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

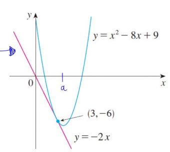
Tangents

เส้นความชัน

- **Definition:** The **tangent line** to $y = f(x)$ at $(a, f(a))$ is the line through $(a, f(a))$ whose slope is equal to $f'(a)$, the derivative of f at a.

- The slope of the tangent line to the graph of f at the point $(a, f(a))$ is also called the **slope of the graph of f at $x = a$**.

อนุพันธ์ = ความชัน



Exercise

- Find the slope of the graph of $f(x) = x^4$ when

(a) $x = -1$

(b) $x = 0$

(c) $x = 1$

$f(x) = x^4$

$f'(x) = 4x^3$

ⓐ $x = -1$

$f'(-1) = 4(-1)^3$

$= -4$

ⓑ $x = 0$

$f'(0) = 4(0)^3$

$= 0$

ⓒ $x = 1$

$f'(1) = 4(1)^3 = 4$

Differentiation Formulas : หาค่าอนุพันธ์

Note: ถ้า d เลขคงที่จะได้ 0 , ถ้า d ตัวแปร x จะได้ 1
ถ้าเป็นสมการปกติก็ d ปกติ

- Derivative of a constant:

$$\frac{d}{dx}(c) = 0$$

- Derivative of x:

$$\frac{d}{dx}(x) = 1$$

- The **Power Rule**: If n is a positive integer, then

$$\boxed{\frac{d}{dx}(x^n) = nx^{n-1}}$$

- The **Constant Multiple Rule**: If c is a constant and f is a differentiable function, then

$$\frac{d}{dx}[cf(x)] = c\frac{d}{dx}f(x)$$

- The **Sum Rule**: If f and g are both differentiable, then

$$\frac{d}{dx}[f(x) + g(x)] = f'(x) + g'(x)$$

- The **Difference Rule**: If f and g are both differentiable, then

$$\frac{d}{dx}[f(x) - g(x)] = f'(x) - g'(x)$$

Example :

1. Find the derivative of $y = x^3 - 4x + 5$.

$$f(x) = x^3 - 4x + 5$$
$$\frac{dy}{dx} = \frac{dx^3}{dx} - 4\overset{=1}{\frac{dx}{dx}} + \overset{0}{\frac{dx5}{dx}}$$
$$= 3x^2 - 4(1)$$
$$= 3x^2 - 4$$

2. Find the derivative of $y = -\frac{x^4}{2} + 3x^3 - 2x$.

$$f(x) = -\frac{x^4}{2} + 3x^3 - 2x$$
$$\frac{dy}{dx} = -\frac{dx^4}{dx\,2} + \frac{3dx^3}{dx} - \frac{2dx}{dx}$$
$$= -\frac{4dx^3}{dx\,2} + \frac{3(3)dx^2}{dx} - \frac{2dx}{dx}$$
$$= -\frac{\cancel{4}x^3}{\cancel{2}} + 9x^2 - 2$$
$$= -2x^3 + 9x^2 - 2$$

- The **Product Rule**: If f and g are both differentiable, then

$$\frac{d}{dx}[\overset{\text{หน้า}}{f(x)}\,\overset{\text{หลัง}}{g(x)}] = f(x)g'(x) + g(x)f'(x)$$

- The **Quotient Rule**: If f and g are differentiable, then

$$\frac{d}{dx}\left[\frac{\overset{\text{บน}}{f(x)}}{\underset{\text{ล่าง}}{g(x)}}\right] = \frac{g(x)f'(x) - f(x)g'(x)}{[g(x)]^2}$$

Example :

1. Find the derivative of $y = (3x - 2x^2)(5 + 4x)$

Note: ถ้าเป็นคูณจะใช้ หน้า d หลัง + หลัง d หน้า

$$f(x) = (3x - 2x^2)(5 + 4x)$$
$$= (3x - 2x^2)\frac{d}{dx}(5 + 4x) + (5 + 4x)\frac{d}{dx}(3x - 2x^2)$$
$$= (3x - 2x^2)(0 + 4(1)) + (5 + 4x)(3 - 4x)$$
$$= (3x - 2x^2)(4) + (5 + 4x)(3 - 4x)$$
$$= (12x - 8x^2) + (15 - 20x + 12x - 16x^2)$$
$$= -24x^2 + 4x + 15 \cancel{x} \qquad 24x - 20$$

2. Find the derivative of $y = \frac{5x-2}{x^2+1}$

Note: ถ้าเป็นหารจะใช้ $\frac{\text{ล่าง d บน + บน d ล่าง}}{(\text{ล่าง})^2}$

$$f(x) = \frac{5x-2}{x^2+1}$$
$$= \frac{(x^2+1)\frac{d}{dx}(5x-2) + (5x-2)\frac{d}{dx}(x^2+1)}{(x^2+1)^2}$$
$$= \frac{(x^2+1)(5-0) + (5x-2)(1+0)}{(x^2+1)^2}$$
$$= \frac{5x^2 + 5 + 5x - 2}{(x^2+1)^2} \quad \Big| \quad = \frac{5x^2 + 5x + 3}{(x^2+1)^2}$$

Chain Rule

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

- or, equivalently,

$$\frac{d}{dx}[f(g(x))] = f'(g(x))g'(x)$$

Example :

1. Find the derivative of $y = (3x - 2x^2)^3$.

$y = (3x - 2x^2)^3$

กำหนด Let $u = (3x - 2x^2)$ then $y = u^3$

$\frac{dy}{dx} = \frac{du^3}{du} \times \frac{du}{du^3}$

$= 3du^2 \times \frac{du}{dx}$

$= 3u^2 \times \left[\frac{d}{dx}(3x - 2x^2)\right]$ 3 4x

$= 3(3x - 2x^2)^2 \times (3 - 4x)$

2. Find the derivative of $y = \left(\frac{3x-1}{x^2+3}\right)^2$.

$y = \left(\frac{3x-1}{x^2+3}\right)^2$

กำหนดให้ Let $u = \left(\frac{3x-1}{x^2+3}\right)$ then $y = u^2$

$\frac{dy}{dx} = \frac{du^2}{du} \cdot \frac{du}{dx}$

$= \frac{2du}{du} \cdot \frac{du}{dx}$

d การหาร

$= 2\left(\frac{3x-1}{x^2+3}\right) \cdot \boxed{\frac{d}{dx}\left(\frac{3x-1}{x^2+3}\right)}$

$= 2\left(\frac{3x-1}{x^2+3}\right) \cdot \frac{(x^2+3)\frac{d}{dx}(3x+1) + (3x-1)\frac{d}{dx}(x^2+3)}{(x^2+3)^2}$ 3 0 1 0

$= 2\left(\frac{3x-1}{x^2+3}\right) \cdot \frac{(x^2+3)(3+0) + (3x-1)(2x+0)}{(x^2+3)^2}$

$= 2\left(\frac{3x-1}{x^2+3}\right) \cdot \frac{(x^2+3)(3) + (3x-1)(2x)}{(x^2+3)^2}$

Maximum and Minimum Values

- **Definition:** Let c be a number in the domain D of a function f. Then $f(c)$ is the

  ทั่วไป

  ค่า function ที่มีค่ามากสุดที่เป็นค่าของ x ที่อยู่ใน D

  - **global maximum** value of f on D if $f(c) \geq f(x)$ for all x in D.
  - **global minimum** value of f on D if $f(c) \leq f(x)$ for all x in D.

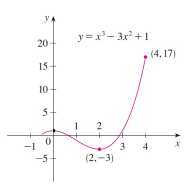
  ค่าต่ำสุดใน domain function

- The **maximum** and **minimum values** of f are called **extreme values** of f.

1. Find the **global maximum** and **minimum** values of the function

① หา critical number

$$f(x) = x^3 - 3x^2 + 1 \text{ and } -\frac{1}{2} \leq x \leq 4$$



$f(x) = x^3 - 3x^2 + 1$ and $\boxed{-\frac{1}{2} \leq x \leq 4}$ → ช่วง

① หา critical $f'(x) = 3x^2 - 2(3x) + 1 \rightarrow \frac{d}{dx}(1) = 0$

$= 3x^2 - 6x$

$3x^2 - 6x = 0$

$3x(x-2) = 0$

$x = 0, 2$ ※

② Therefore, the critical numbers of f are $x = 0$ and $x = 2$

The value of f at the critical numbers are

$f(0) = f(x)$
$= (0)^3 - 3(0)^2 + 1$
$= 1$

$f(2) = f(x)$
$= (2)^3 - 3(2)^2 + 1$
$= 8 - 12 + 1$
$= -3$ → global min

③ หาค่า global

The values of f at the endpoints of the interval are $-\frac{1}{2} \leq x \leq 4$

$f(x) = x^3 - 3x^2 + 1$

$f(-\frac{1}{2}) = (-\frac{1}{2})^3 - 3(-\frac{1}{2})^2 + 1$
$= -\frac{1}{8} - \frac{3}{4} + 1$

$\frac{-1-6+8}{8}$

$\frac{-7+8}{8} = \frac{1}{8}$ ※

$f(4) = (4)^3 - 3(4)^2 + 1$
$= 64 - 48 + 1$
$= 17$ → global max

∴ global maximum value is 17

global minimum value is −3