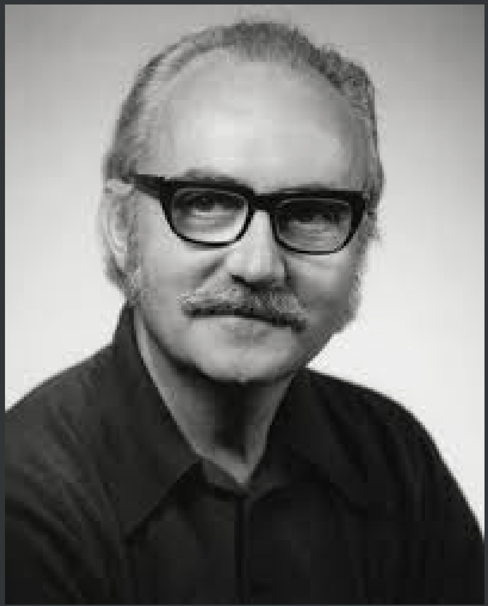
Discrete Choice

All models are wrong, some are useful.

--- George Box



Question:

How do machines recognize hand-written digits?

What brand is my smartphone?

Apple

Xiaomi

Huawei

OPPO

I am using a Xiaomi phone.

What is the brand of the HKU president's car?

TESLA

Mercedes-Benz

BMW

Audi

港大校長座駕換車　張翔指定寶馬i7　205萬元豁免招標惹爭議

撰文：勞顯亮

出版：2023-09-12 07:00　更新：2023-09-12 13:37

基于环保考虑，大学的校园设施管理部门建议以纯电动车代替燃油车。在比较了市面上多款电动车及混合动力型车的资料后，该部门与校长室挑选了以上两款，安排校长及两个部门的同事于同一天参与试车，并归纳了各人的反馈意見。校园设施管理部门同时比较了两款车的性能表现、车厢容量、外观、价格，亦考虑了大学的需要。

Daniel McFadden’s developed a model to understand our choices. His model became so popular, and he won the Nobel Prize in Economics in 2000 for “his development of theory and methods for analyzing discrete choice.”

Modelling Consumer Choice

Human-beings always need to make choices, from your marriage choice to buying a bottle of milk.

While individuals can make choices in their own ways, as consumer analysts, we do want to understand how consumers make their choices.

Imaging that you are a bank manager.

You want to understand how consumers choose between different credit card companies when applying for credit cards. In this way, you can understand who are really your potential clients, and you can target on these consumers better.

Your data is as follows...

For each consumer, you know his or her demographics (e.g., gender, age), occupation, income, geographic location, credit histories, etc. These are your independent variables.

You also know which credit card they applied to, e.g., Citibank, HSBC, BOC, American Express, ... or none of the above. This is your dependent variable.

You task: Building a model that predicts the dependent variable using your independent variables.

What would you do?

Let us start with something simpler.

Now, you want to predict whether or not a consumer applies for your company's credit card. Here, the dependent variable Y_i is YES or NO. For simplicity, let $Y_i = 1$ for YES and $Y_i = 0$ for NO.

For each individual, the independent variables again include demographics, occupation, income, location, etc. We use X_i to denote the independent variables.

Our task: Predict Y_i using X_i.

What should you do?

Our task: Predict Y_i using X_i, where $Y_i \in \{0, 1\}$.

Question: Can we use linear regression to analyze the relationship between Y_i and X_i, that is, we use the following linear model:

$$Y_i = \alpha + \beta X_i$$

Issues with linear regression

Suppose that your regression result is:

$$Y_i = 0.4 + 0.1 \times Age_i + 0.2 \times Female_i$$

Suppose that a person's age is 25 and gender is male, you predict that his $Y_i = 0.65$, that is, the person is likely to buy from you.

Issues with linear regression

Suppose that your regression result is:

$$Y_i = 0.4 + 0.1 \times Age_i + 0.3 \times Female_i$$

Suppose that another person's age is 40 and gender is female, you predict that her $Y_i = 1.1$.

How would you interpret this result? Will she apply for your credit card 1.1 times? It does not make any sense!

What should we do?

Instead of predicting the value of Y_i directly, we can predict the probability that Y_i is equal to 1, i.e., we want to predict $\Pr[Y_i = 1]$.

How to do that? We want to find out a function f such that

$$\Pr[Y_i = 1] \approx f(X_i)$$

Next, we will look for such a function f.

What should we do?

How to do that? We want to find out a function f such that

$$\Pr[Y_i = 1] \approx f(X_i)$$

Here, we need to impose some restrictions on the function f:

1. $f(X) \geq 0$ for all X: probabilities are nonnegative.
2. $f(X) \leq 1$ for all X: probabilities are no more than 100%.
3. $f(X)$ is either increasing or decreasing with X.

What should we do?

$$\Pr[Y_i = 1] \approx f(X_i)$$

Here, we need to impose some restrictions on function f:

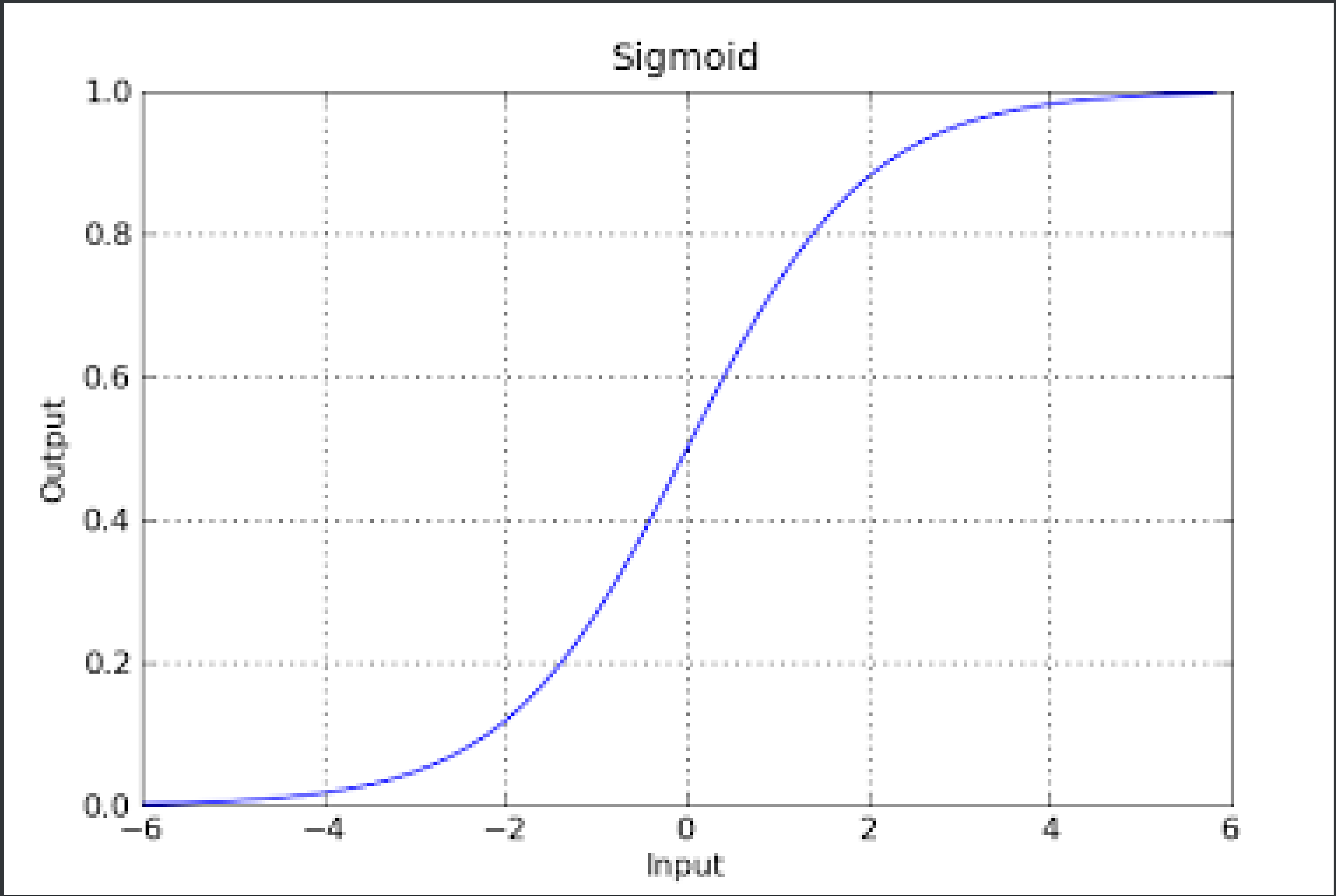
1. $f(X) \geq 0$ for all X: probabilities are nonnegative.
2. $f(X) \leq 1$ for all X: probabilities are no more than 100%.
3. $f(X)$ is either increasing or decreasing with X.

Can you propose such a function f? Any ideas?

$$f(X) = \frac{\exp(\alpha + \beta X)}{1 + \exp(\alpha + \beta X)}$$

Note: $\exp(x) = e^x$ is the exponential function.

When $\beta > 0$, $f(X)$ increases with X; when $\beta < 0$, $f(X)$ decreases with X.



The logistic function

A video explaining logistic function

Our task

We already know the values X_i and Y_i for each individual i. We would like to find the values of α and β to approximate the relationship between X_i and Y_i:

$$\Pr(Y_i = 1) \approx \frac{\exp(\alpha + \beta X_i)}{1 + \exp(\alpha + \beta X_i)}$$

This is done via maximum likelihood estimation. Check here if you want to know more details.

As an illustration, we first load the following dataset in R.

```r
library(readr)
mydata <- read.csv("https://ximarketing.github.io/data/banking.csv")
head(mydata)
```

The data reads as follows:

	age	job	previous	success
1	44	blue-collar	0	0
2	53	technician	0	0
3	28	management	2	1
4	39	services	0	0
5	55	retired	1	1
6	30	management	0	0

	age	job	previous	success
1	44	blue-collar	0	0
2	53	technician	0	0
3	28	management	2	1
4	39	services	0	0
5	55	retired	1	1
6	30	management	0	0

The data is about the outcome of a marketing campaign in a Portuguese banking that promotes a term deposit to their clients. Success denotes the final outcome of the campaign (1 = success, 0 = failure).

- Job includes admin, blue-collar, entrepreneur, housemaid, management, retired, self-employed, services, student, technician, unemployed, unknown.
- Previous denotes the number of previous interactions with the client.

m deposit

```r
result <- glm(success ~ age + factor(job) + previous, data
              = mydata, family = "binomial")
summary(result)
```

Next, we build up a logistic regression model using success to be the dependent variable, independent variables include age, job, and number of previous contacts.

Note that because “job” is not a number, we treat it as a fixed effect by enclosing it within a factor bracket.

```
Coefficients:
                           Estimate Std. Error z value Pr(>|z|)    
(Intercept)               -2.217305   0.074546 -29.744  < 2e-16 ***
age                        0.001890   0.001776   1.064 0.287149    
factor(job)blue-collar    -0.625683   0.051213 -12.217  < 2e-16 ***
factor(job)entrepreneur   -0.416913   0.099843  -4.176 2.97e-05 ***
factor(job)housemaid      -0.257722   0.109806  -2.347 0.018922 *  
factor(job)management     -0.174238   0.067648  -2.576 0.010005 *  
factor(job)retired         0.667628   0.078456   8.510  < 2e-16 ***
factor(job)self-employed -0.188631   0.093062  -2.027 0.042670 *  
factor(job)services       -0.487867   0.066121  -7.378 1.60e-13 ***
factor(job)student         0.879372   0.086922  10.117  < 2e-16 ***
factor(job)technician     -0.168579   0.050048  -3.368 0.000756 ***
factor(job)unemployed      0.093801   0.097300   0.964 0.335027    
factor(job)unknown        -0.150583   0.181433  -0.830 0.406558    
previous                   0.879022   0.024831  35.401  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

How to interpret these results?

We look at the estimates and the p-value (significance).

Age is not significant; it means whether a client accepts your promotion has little to do with his or her age.

Previous is significant and positive, meaning that getting a deal is easier when you have more previous interactions with the client.

Lastly, which types of jobs are more likely to accept your promotion? Retired and student. On the other hand, blue-collar, services, and entrepreneurs are unlikely to be convinced.

Probit regression

Probit Regression

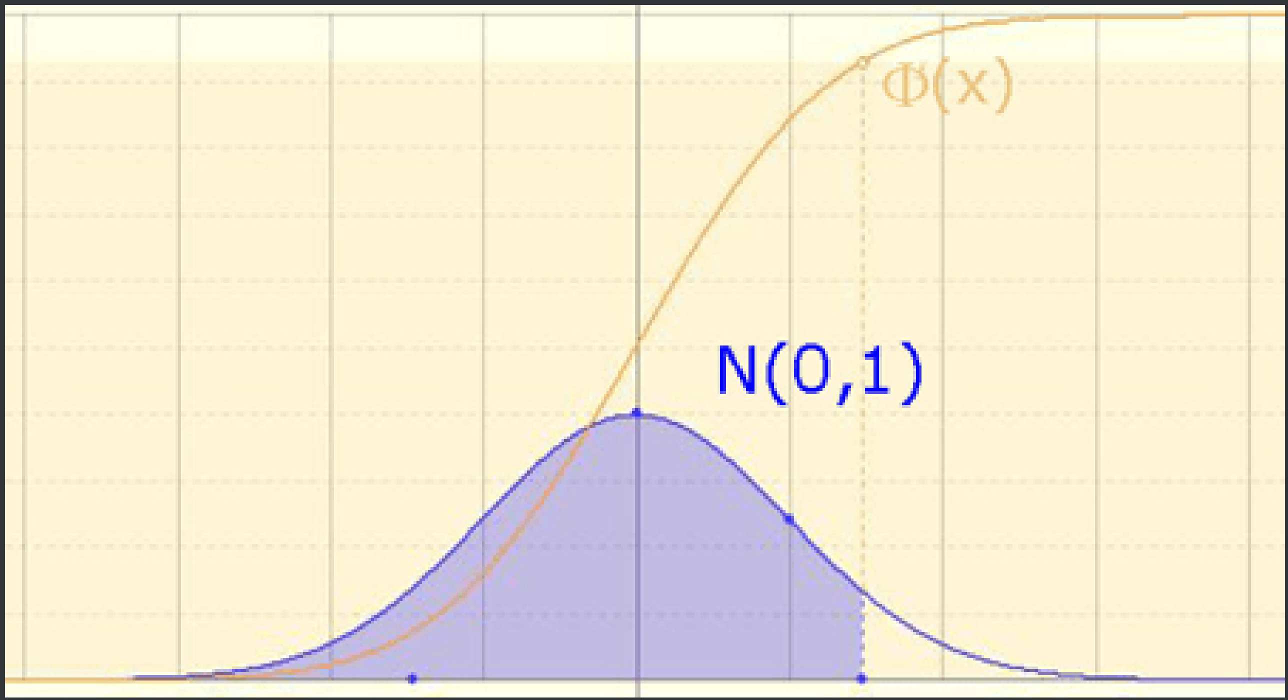
In logistic regression, we adopt the logistic function to estimate $\Pr\left[Y = 1 \,\middle|\, X\right]$, which satisfies the properties that we listed. However, the logistic function is not the only function that satisfies those properties. Now, we introduce another function that can also make predictions about binary outcomes.

Probit Regression

Here, we use the cumulative distribution function of the standard normal distribution. Mathematically, suppose that $v \sim N(0, 1)$ is a standard normal random variable, then we can define the cumulative distribution function Φ as

$$\Phi(z) = \Pr[v \leq z].$$

Probit Regression



Probit Regression

```r
library(readr)
mydata <- read.csv("https://ximarketing.github.io/data/banking.csv")
head(mydata)
probit   <- glm(success ~ age + factor(job) + previous, data
               = mydata, family = binomial(link = "probit"))
summary(probit)
```

	Dependent variable:	
	success	
	logistic	*probit*
	(1)	(2)
age	0.002	0.0004
	(0.002)	(0.001)
factor(job)blue-collar	-0.626***	-0.312***
	(0.051)	(0.026)
factor(job)entrepreneur	-0.417***	-0.205***
	(0.100)	(0.050)
factor(job)housemaid	-0.258**	-0.138**
	(0.110)	(0.057)
factor(job)management	-0.174**	-0.090**
	(0.068)	(0.035)
factor(job)retired	0.668***	0.391***
	(0.078)	(0.044)
factor(job)self-employed	-0.189**	-0.093*
	(0.093)	(0.048)
factor(job)services	-0.488***	-0.247***
	(0.066)	(0.033)
factor(job)student	0.879***	0.497***
	(0.087)	(0.050)
factor(job)technician	-0.169***	-0.092***
	(0.050)	(0.026)
factor(job)unemployed	0.094	0.046
	(0.097)	(0.053)
factor(job)unknown	-0.151	-0.084
	(0.181)	(0.095)
previous	0.879***	0.494***
	(0.025)	(0.014)
Constant	-2.217***	-1.273***
	(0.075)	(0.039)
Observations	41,188	41,188
Log Likelihood	-13,462.880	-13,460.430
Akaike Inf. Crit.	26,953.770	26,948.860
Note:	*p<0.1; **p<0.05; ***p<0.01	

Logistic vs. Probit

Question: Which one makes more sense?

Logistics vs. Probit

They are similar models that yield similar (though not identical) inferences.

- Logistic regression is more popular in healthcare.
- Probit regression is more popular in political science.

But in most situations, it does not matter which method you choose to go with. Working with either will be fine.

The next question: What should we do when consumers have more than two choices?

More specifically, let us consider the following problem.

Each consumer i has his or her own information, which is measured by the independent variable X_i. The dependent variable is a choice made by the consumer, $Y_i \in \{A, B, \ldots\}$.

More specifically, let us consider the following problem.

Each consumer i has his or her own information, which is measured by the independent variable X_i. The dependent variable is a choice made by the consumer, $Y_i \in \{A, B, \ldots\}$.

Idea: Instead of predicting Y_i directly, we predict the probability $\Pr[Y_i = A], \Pr[Y_i = B], \ldots$

Suppose that consumers have three choices, A, B, C.

Now, given X_i, we would like to come up with three functions $f_A(X_i)$, $f_B(X_i)$ and $f_C(X_i)$, such that

$$\begin{aligned}
\Pr[Y_i = A] &\approx f_A(X_i), \\
\Pr[Y_i = B] &\approx f_B(X_i), \\
\Pr[Y_i = C] &\approx f_C(X_i).
\end{aligned}$$

As before, we place a few restrictions on these functions:

1. The probabilities must be nonnegative, i.e., $f_j(X_i) \geq 0$
2. Probabilities cannot exceed 1, i.e., $f_j(X_i) \leq 1$
3. Probabilities are monotone with X_i
4. Now, we have a new constraint: all the probabilities must add up to 100%, i.e.,

$$f_A(X_i) + f_B(X_i) + f_C(X_i) = 1.$$

Any ideas for the functions?

$$f_A(X_i) = \frac{\exp(\alpha_A + \beta_A X_i)}{\exp(\alpha_A + \beta_A X_i) + \exp(\alpha_B + \beta_B X_i) + \exp(\alpha_C + \beta_C X_i)}$$

$$f_B(X_i) = \frac{\exp(\alpha_B + \beta_B X_i)}{\exp(\alpha_A + \beta_A X_i) + \exp(\alpha_B + \beta_B X_i) + \exp(\alpha_C + \beta_C X_i)}$$

$$f_C(X_i) = \frac{\exp(\alpha_C + \beta_C X_i)}{\exp(\alpha_A + \beta_A X_i) + \exp(\alpha_B + \beta_B X_i) + \exp(\alpha_C + \beta_C X_i)}$$

They satisfy all the constraints!

We need to estimate the values of α's and β's.

```r
library(foreign)
library(nnet)
library(stargazer)
```

We install and load several packages for multinomial logit regression.

We first load the data from the Internet.

```r
mydata <-
read.csv("https://ximarketing.github.io/data/multinomial_route_choice.csv")
head(mydata)
```

Here is the data...

	Choice	Flow	Distance	Seat_belt	Passengers	Age	Male	Income	Fuel_efficiency
1	Arterial	460	48	0	0	2	0	1	28
2	Rural	440	44	0	0	2	0	1	28
3	Freeway	130	61	0	0	2	0	1	28
4	Arterial	595	59	1	0	2	1	2	27
5	Rural	515	70	1	0	2	1	2	27
6	Freeway	340	87	1	0	2	1	2	27

Here, we want to predict how individuals choose the route when driving. The dependent variable is the chosen route, which can be arterial, rural, and freeway.

The independent variables include the followings:

Flow: A measure of traffic flow (how busy the traffic is).

Distance: The distance of the planned trip.

Seat_belt: whether the driver wears seat belt.

Passengers: Number of passengers carried.

Age: Age group of the driver.

Male: Whether the driver is male or not.

Income: Income level of the driver.

Fuel_efficiency: Fuel efficiency level of the vehicle.

We use the multinom function to perform multinomial logit regression:

```r
result <- multinom(formula = Choice ~ Flow + Distance +
                             Seat_belt + Passengers + Age + Male +
                             Income + Fuel_efficiency, data = mydata)
result
```

Oh, the results do not read nicely...

```
Coefficients:
        (Intercept)         Flow   Distance  Seat_belt Passengers         Age        Male
Freeway   13.673284 -0.049143703  0.1362782 -0.8924558  0.4775758  0.17728498  0.06331663
Rural      7.558223 -0.008436186 -0.0455514 -0.3451560  0.1436887 -0.06181751 -0.04244764
            Income Fuel_efficiency
Freeway -0.5430466     -0.06321059
Rural    0.1319585     -0.01778424
```

No worries, let's try the stargazer function.

```
stargazer(result, type="html", out="result.html")
```

Now, our results are nicely summarized in the table on the right-hand side:

What does it mean?

	Dependent variable:	
	Freeway	Rural
	(1)	(2)
Flow	-0.049***	-0.008***
	(0.006)	(0.001)
Distance	0.136***	-0.046***
	(0.031)	(0.014)
Seat_belt	-0.892	-0.345
	(0.663)	(0.319)
Passengers	0.478	0.144
	(0.454)	(0.275)
Age	0.177	-0.062
	(0.310)	(0.157)
Male	0.063	-0.042
	(0.638)	(0.302)
Income	-0.543	0.132
	(0.379)	(0.144)
Fuel_efficiency	-0.063	-0.018
	(0.068)	(0.038)
Constant	13.673***	7.558***
	(0.158)	(1.390)
Akaike Inf. Crit.	419.424	419.424
Note:		*p<0.1; **p<0.05; ***p<0.01

	Dependent variable:	
	Freeway	Rural
	(1)	(2)
Flow	-0.049***	-0.008***
	(0.006)	(0.001)
Distance	0.136***	-0.046***
	(0.031)	(0.014)
Seat_belt	-0.892	-0.345
	(0.663)	(0.319)
Passengers	0.478	0.144
	(0.454)	(0.275)
Age	0.177	-0.062
	(0.310)	(0.157)
Male	0.063	-0.042
	(0.638)	(0.302)
Income	-0.543	0.132
	(0.379)	(0.144)
Fuel_efficiency	-0.063	-0.018
	(0.068)	(0.038)
Constant	13.673***	7.558***
	(0.158)	(1.390)
Akaike Inf. Crit.	419.424	419.424
Note:	*p<0.1; **p<0.05; ***p<0.01	

Here, we take arterial as the benchmark and compare other routes against it. Alternatively, you can view the parameters for arterial to be equal to zero.

Flow: When there is a high flow, drivers are very less likely to choose freeway, and a bit less likely to choose rural compared with arterial.

Distance: When distance is long, drivers are more likely to choose freeway and less likely to choose rural route...

The complete code is here:

```r
library(foreign)
library(nnet)
library(stargazer)
mydata <-
read.csv("https://ximarketing.github.io/data/multinomial_route_choice.csv"
)
head(mydata)
result <- multinom(formula = Choice ~ Flow + Distance +
                     Seat_belt + Passengers + Age + Male +
                     Income + Fuel_efficiency, data = mydata)
result
stargazer(result, type="html", out="result.html")
```

Back to the Question:

How do machines recognize hand-written digits?

Back to the Question:

How do machines recognize hand-written digits?

Absolutely, there are many sophisticated algorithms for handwriting recognition such as convolutional neural networks. But in the early stage, scientists just use the multinomial logit model to perform the task.

Input: Handwriting in pixels.

Output: $Y_i \in \{0, 1, \ldots, 9\}$

Conditional Logit Model

In multinomial logit model, a person chooses among a few alternatives. The decision hinges on the decision maker's personal features, not the features of the alternatives. In our previous example, the route decision hinges on features such as distance, age, which are constant across all alternatives.

In conditional logit model, a person chooses among a few alternatives. The decision hinges on the alternatives' features, not the feature of the individuals.

Example:

Consumers choose among three computers, A, B, and C.

1. If the choices are based on consumers' age, gender, education etc, then we use the multinomial logit model.
2. If the choices are based on the price, quality of the computers, then we use the conditional logit model.

```r
library(survival)
library(stargazer)
mydata = read.csv("https://ximarketing.github.io/data/conjoint.csv")
head(mydata)
```

	id	price	storage	ram	cpu	choice
1	1	400	512	4	3.6	1
2	1	400	256	8	2.8	0
3	1	300	128	4	5.0	0
4	2	500	256	2	5.0	0
5	2	300	512	8	2.8	0
6	2	400	512	4	3.6	1

	id	price	storage	ram	cpu	choice
1	1	400	512	4	3.6	1
2	1	400	256	8	2.8	0
3	1	300	128	4	5.0	0
4	2	500	256	2	5.0	0
5	2	300	512	8	2.8	0
6	2	400	512	4	3.6	1

Consumer 1 (id = 1) chooses between three computers:

1. Price = 400, Storage = 512 GB, RAM = 4 GB, CPU = 3.6 GHz
2. Price = 400, Storage = 256 GB, RAM = 8 GB, CPU = 2.8 GHz
3. Price = 300, Storage = 128 GB, RAM = 4 GB, CPU = 5.0 GHz

And this consumer chooses the first computer (choice = 1)

```r
result<-clogit(choice ~ price + cpu +
                 ram + storage + strata(id), data=mydata)
summary(result)
```

```
             coef  exp(coef)   se(coef)      z Pr(>|z|)    
price  -0.0038226  0.9961847  0.0004281 -8.929   <2e-16 ***
cpu     0.4974295  1.6444886  0.0378409 13.145   <2e-16 ***
ram     0.1486753  1.1602962  0.0070257 21.162   <2e-16 ***
storage 0.0055173  1.0055325  0.0002284 24.159   <2e-16 ***
```

```r
stargazer(result, type="html", out="result.html")
```

	Dependent variable:
	choice
price	-0.003***
	(0.0004)
cpu	0.366***
	(0.027)
ram	0.138***
	(0.007)
storage	0.005***
	(0.0002)
Observations	6,000
R^2	0.198
Max. Possible R^2	0.519
Log Likelihood	-1,537.106
Wald Test	790.650*** (df = 4)
LR Test	1,320.236*** (df = 4)
Score (Logrank) Test	1,155.273*** (df = 4)
Note:	*p<0.1; **p<0.05; ***p<0.01

When price increases, the computer is less likely to be chosen; when CPU, RAM or Storage increases, the computer is more likely to be chosen.

	coef	exp(coef)	se(coef)	z	Pr(>|z|)	
price	-0.0038226	0.9961847	0.0004281	-8.929	<2e-16	***
cpu	0.4974295	1.6444886	0.0378409	13.145	<2e-16	***
ram	0.1486753	1.1602962	0.0070257	21.162	<2e-16	***
storage	0.0055173	1.0055325	0.0002284	24.159	<2e-16	***

The coefficient for price is -0.0038 and the coefficient for RAM is 0.1486. Because $0.1486/0.0038 = 38.8$, it suggests that a 1GB increase in RAM is equivalent to a \$38.8 decrease in price. Or put differently, 1 GB RAM is worth \$38.8 to an average consumer.

The complete code is here:

```r
library(survival)
library(stargazer)
mydata = read.csv("https://ximarketing.github.io/data/conjoint.csv")
head(mydata)
result<-clogit(choice ~ price + cpu +
                 ram + storage + strata(id), data=mydata)
summary(result)
stargazer(result, type="html", out="result.html")
```

Predicting Market Share

Suppose that there are two PCs available in the market:

- (1) Price = 400, CPU = 3.6 GHz, RAM = 4 GB, Storage = 512 GB
- (2) Price = 280, CPU = 3.2 GHz, RAM = 4 GB, Storage = 256 GB

We can use our regression results to predict their market share, following the formula of conditional logit.

```r
library(survival)
library(stargazer)
mydata = read.csv("https://ximarketing.github.io/data/conjoint.csv")
head(mydata)
result<-clogit(choice ~ price + cpu +
                 ram + storage + strata(id), data=mydata)

coef_price <- coef(result)["price"]
coef_cpu <- coef(result)["cpu"]
coef_ram <- coef(result)["ram"]
coef_storage <- coef(result)["storage"]

price1 <- 400; cpu1 <- 3.6; ram1 <- 4; storage1 <- 512
price2 <- 280; cpu2 <- 3.2; ram2 <- 4; storage2 <- 256

d1 <- exp(price1 * coef_price + cpu1 * coef_cpu + ram1 * coef_ram +
storage1 * coef_storage)
d2 <- exp(price2 * coef_price + cpu2 * coef_cpu + ram2 * coef_ram +
storage2 * coef_storage)

s1 <- d1/(d1+d2)
s2 <- d2/(d1+d2)
print(c(s1, s2))
```

Other Models

There are also many other models beyond ones we discuss in class:

- If your dependent variable is the number of units (e.g., X bottles of milk; Y individuals...), you can use Poisson regression.
- If your dependent variable is censored (e.g., you only observe those whose income is greater than 100K), you can use Tobit model.

https://www.youtube.com/embed/i8tjLQUPc8Y?enablejsapi=1