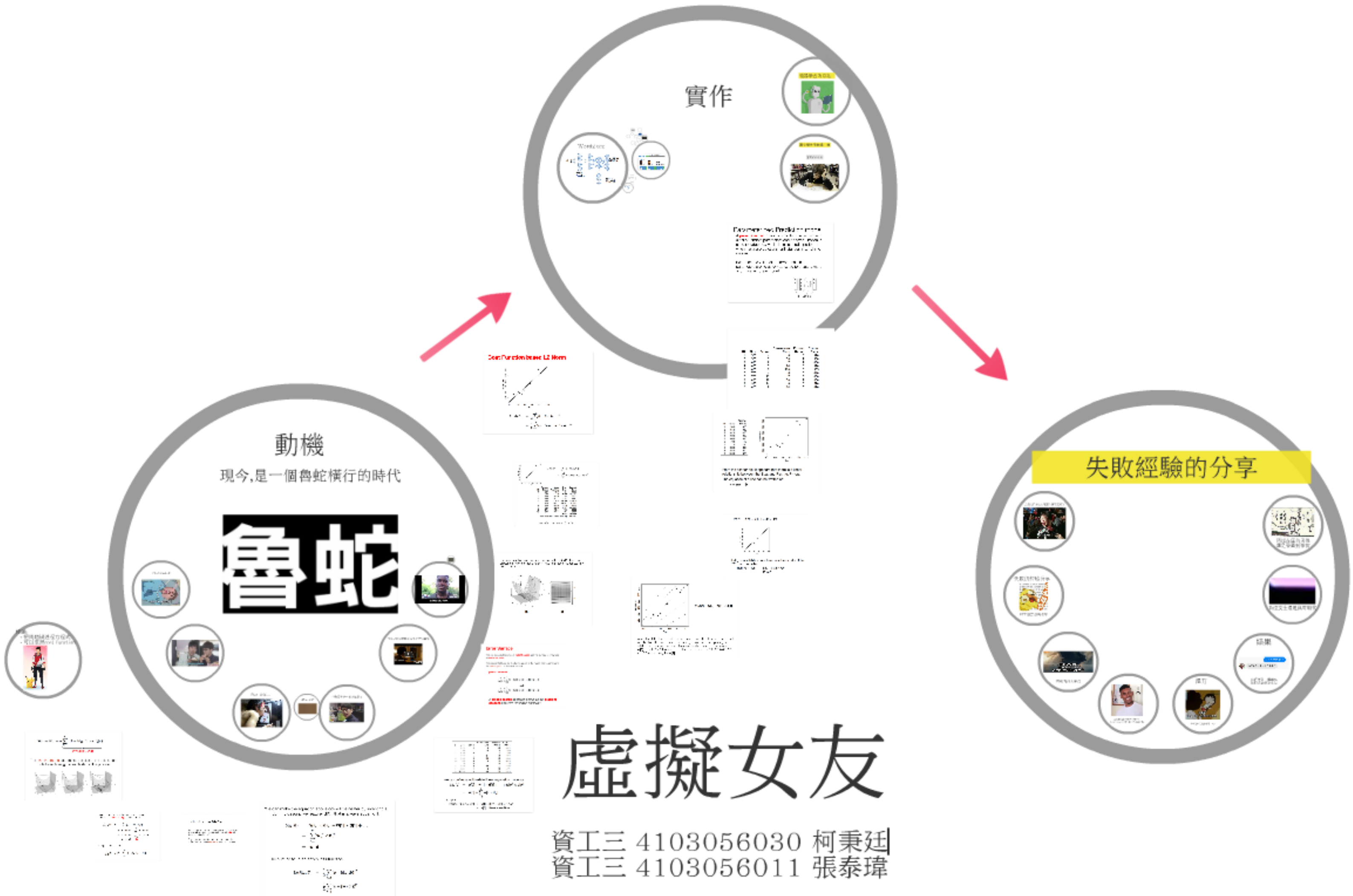
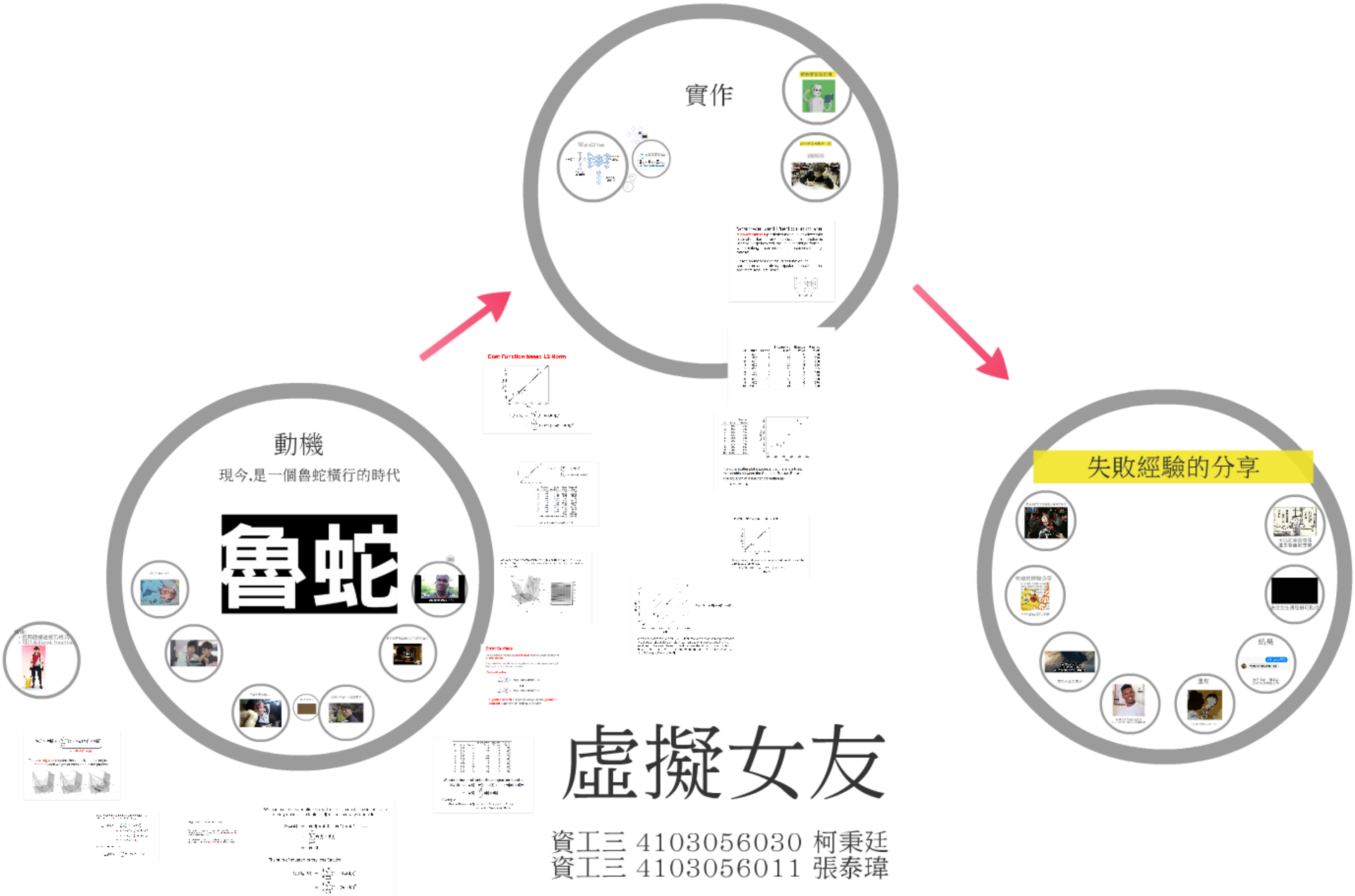



# 虛擬女友

資工三 4103056030 柯秉廷
資工三 4103056011 張泰瑋

失敗的理由
0. 雖然我們失敗了，但是還是有些經
驗可以跟大家分享:
0. 不要前一天再開始做
1..PTT講話太沒水準
2. 還是沒有很好的語料去訓練機器人
3. 訓練過程中，也沒有一個很好的
benchmark (指標)，去衡量機器人回
一句話到底好不好

結論:
1.算是失敗收場，只能得到一個講話
很賤的女朋友
2.所以在座女生可以放心，你們暫時
不會被人工智慧取代QQ


實作
動機
現今,是一個魯蛇橫行的時代
魯蛇
失敗經驗的分享
虛擬女友
資工三 4103056030 柯秉廷
資工三 4103056011 張泰瑋

# 動機

## 現今,是一個魯蛇橫行的時代

# 魯蛇

日本已經有虛擬女友了,但要8萬元

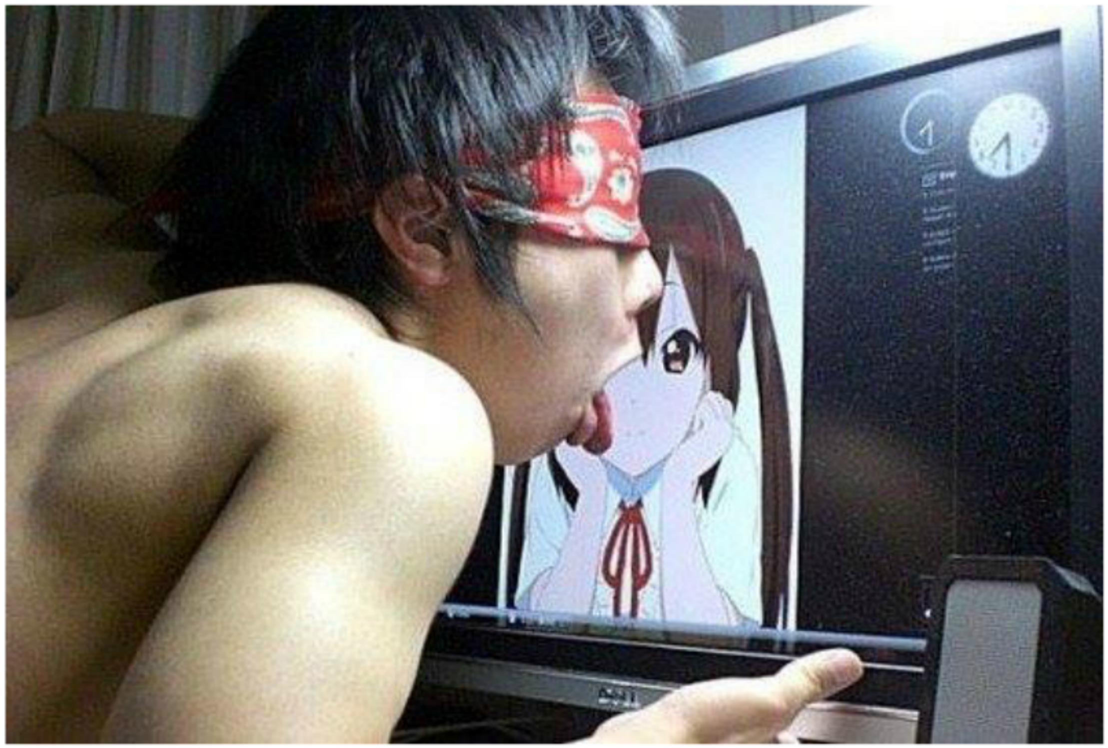
只能在電腦前....

但現實生活中我還是肥宅

原本只有偶爾被閃瞎


我今天带的也是爱妻便当
俺は今日も愛妻弁当

只能在電腦前....

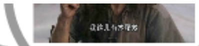


# 剛看完月薪嬌妻

薪嬌妻


# 但現實生活中我還是肥宅

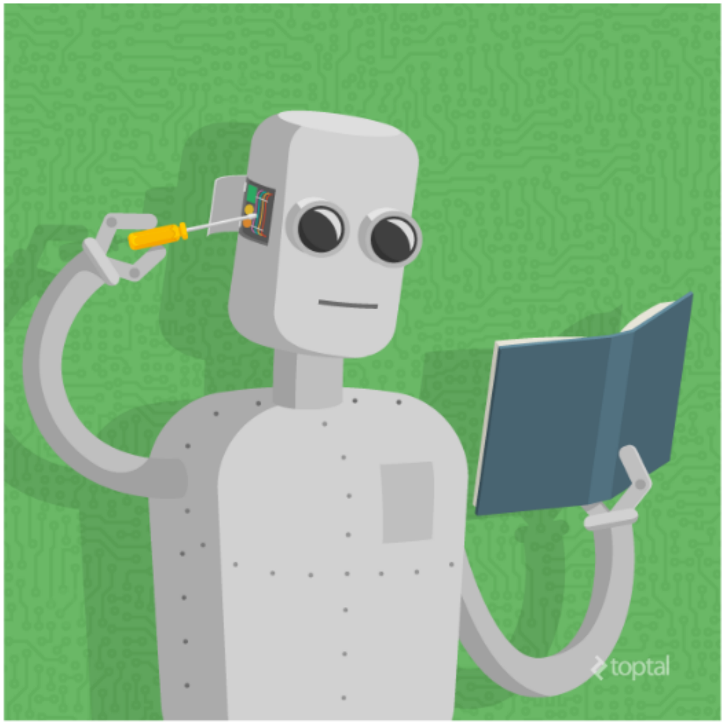
日本已經有虛擬女友了,但要8萬元




影片翻攝
SHEWOLF
怎麼你突然哭起來了？？


如来神掌
我这儿有本秘笈

# 機器學習的原理

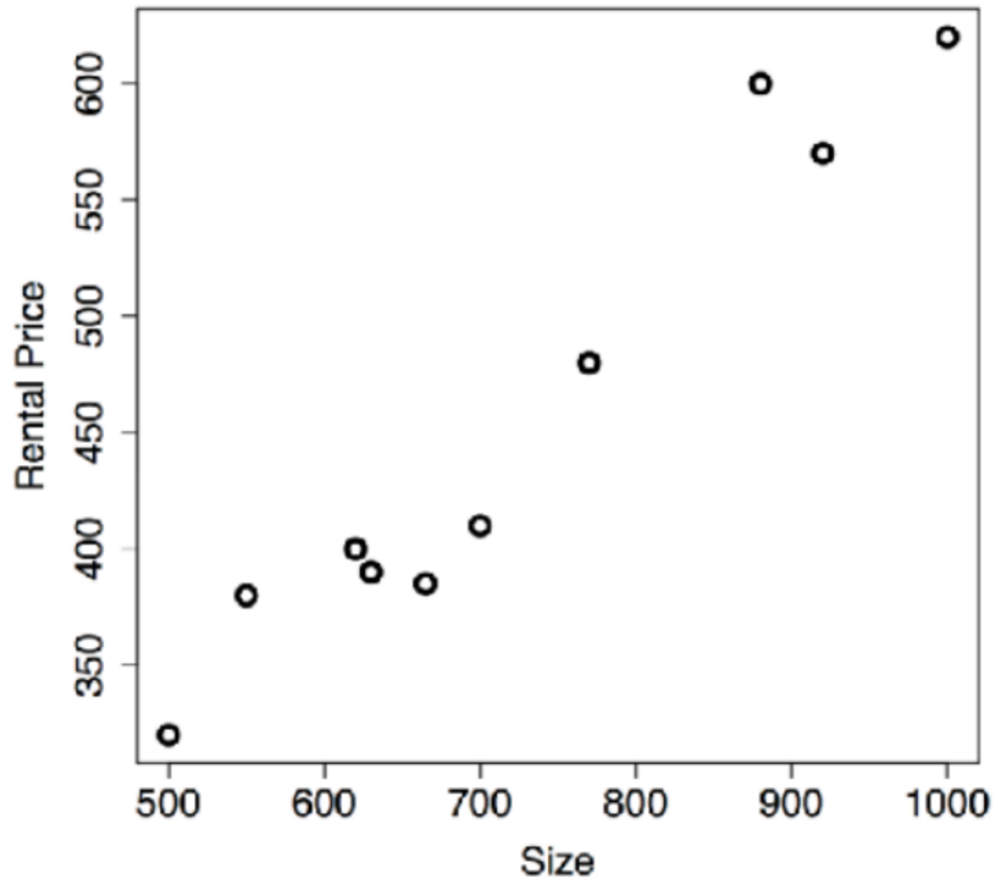


跟小學生考數學一樣

2X+Y=10

| ID | Size | Floor | Broadband Rate | Energy Rating | Rental Price |
|---|---|---|---|---|---|
| 1 | 500 | 4 | 8 | C | 320 |
| 2 | 550 | 7 | 50 | A | 380 |
| 3 | 620 | 9 | 7 | A | 400 |
| 4 | 630 | 5 | 24 | B | 390 |
| 5 | 665 | 8 | 100 | C | 385 |
| 6 | 700 | 4 | 8 | B | 410 |
| 7 | 770 | 10 | 7 | B | 480 |
| 8 | 880 | 12 | 50 | A | 600 |
| 9 | 920 | 14 | 8 | C | 570 |
| 10 | 1,000 | 9 | 24 | B | 620 |

| ID | Size | Rental Price |
|---|---|---|
| 1 | 500 | 320 |
| 2 | 550 | 380 |
| 3 | 620 | 400 |
| 4 | 630 | 390 |
| 5 | 665 | 385 |
| 6 | 700 | 410 |
| 7 | 770 | 480 |
| 8 | 880 | 600 |
| 9 | 920 | 570 |
| 10 | 1,000 | 620 |

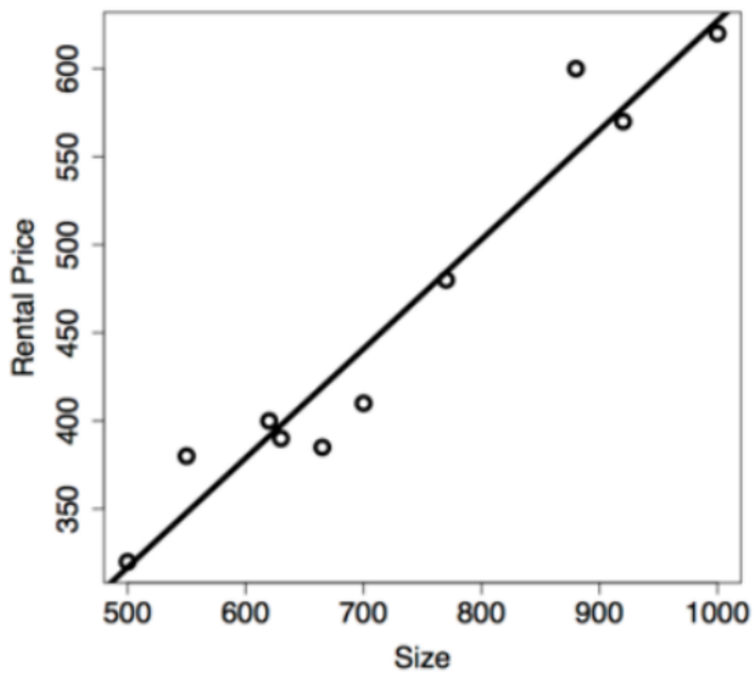



From the scatter plot it appears that there is a linear relationship between the SIZE and RENTAL PRICE.

The equation of a line can be written as:

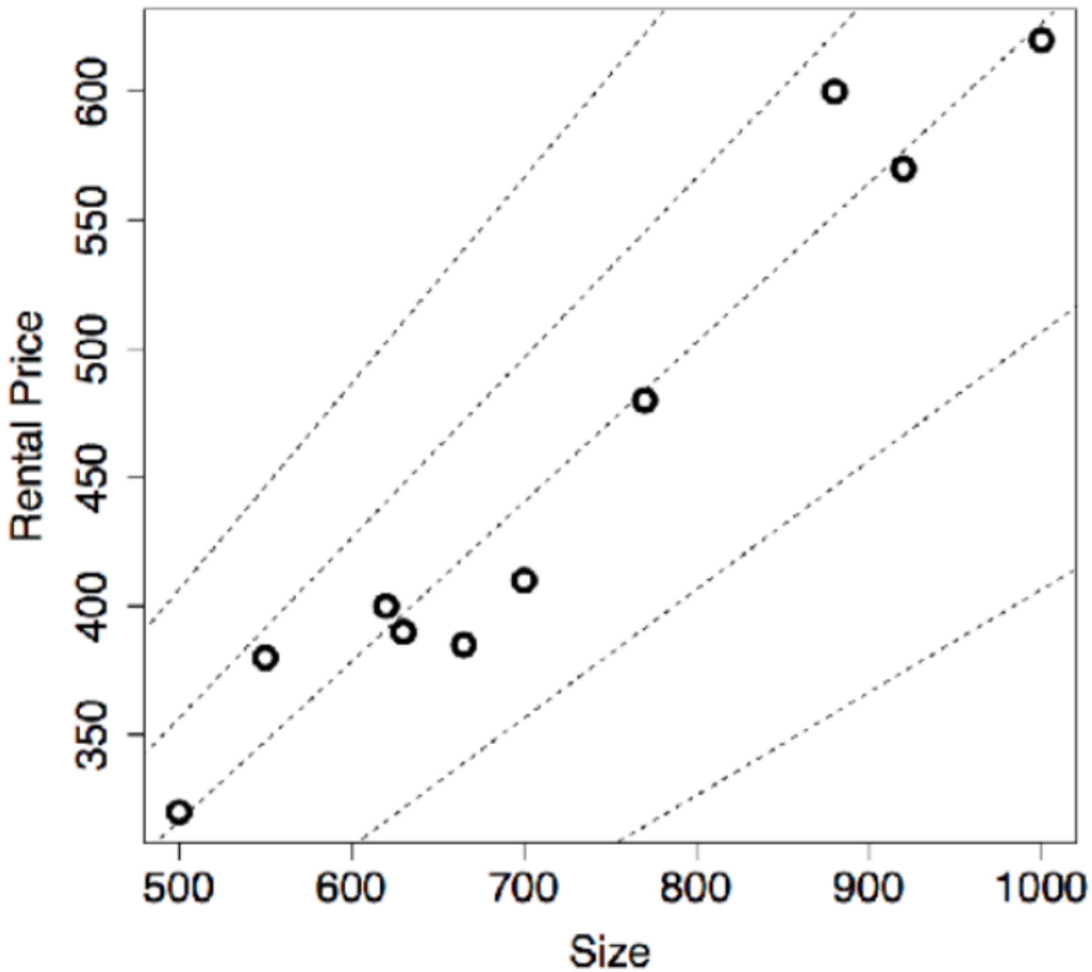
$$y = mx + b$$

$$\text{RENTAL PRICE} = 6.47 + 0.62 \times \text{SIZE}$$




Using this model determine the expected rental price of the 730 square foot office:

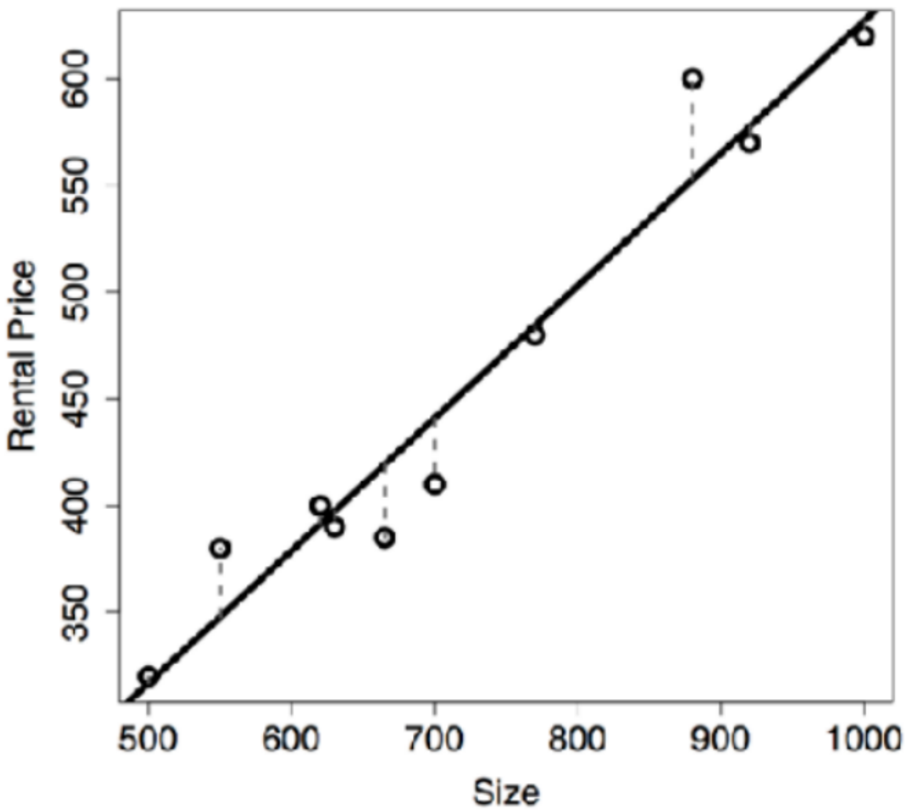
$$\begin{aligned}\text{RENTAL PRICE} &= 6.47 + 0.62 \times 730 \\ &= 459.07\end{aligned}$$



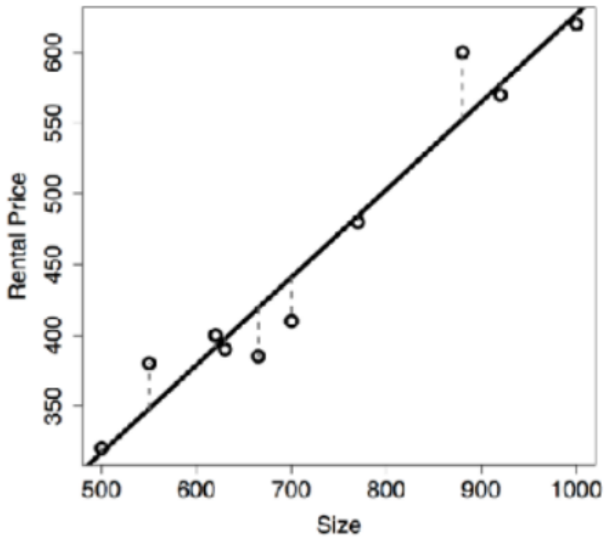

$$\mathbb{M}_{\mathbf{w}}(d) = \mathbf{w}[0] + \mathbf{w}[1] \times \mathbf{d}[1]$$

A scatter plot of the SIZE and RENTAL PRICE features from the office rentals dataset. A collection of possible simple linear regression models capturing the relationship between these two features are also shown. For all models **w[0]** is set to 6.47. From top to bottom the models use 0.4, 0.5, 0.62, 0.7 and 0.8 respectively for **w[1]**.

# Cost Function based L2 Norm




$$
\begin{aligned}
L_2(\mathbb{M}_{\mathbf{w}}, \mathcal{D}) &= \frac{1}{2}\sum_{i=1}^{n}\left(t_i - \mathbb{M}_{\mathbf{w}}(\mathbf{d}_i[1])\right)^2 \\
&= \frac{1}{2}\sum_{i=1}^{n}\left(t_i - (\mathbf{w}[0] + \mathbf{w}[1] \times \mathbf{d}_i[1])\right)^2
\end{aligned}
$$



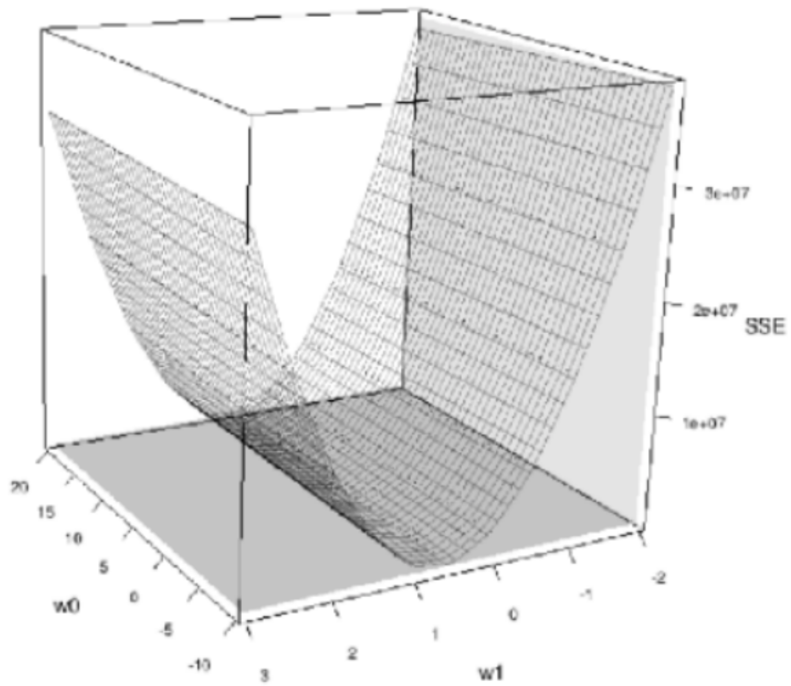

$$L_2(\mathbb{M}_{\mathbf{w}}, \mathcal{D}) = \frac{1}{2}\sum_{i=1}^{n}(t_i - \mathbb{M}_{\mathbf{w}}(\mathbf{d}_i[1]))^2$$

$$= \frac{1}{2}\sum_{i=1}^{n}(t_i - (\mathbf{w}[0] + \mathbf{w}[1] \times \mathbf{d}_i[1]))^2$$

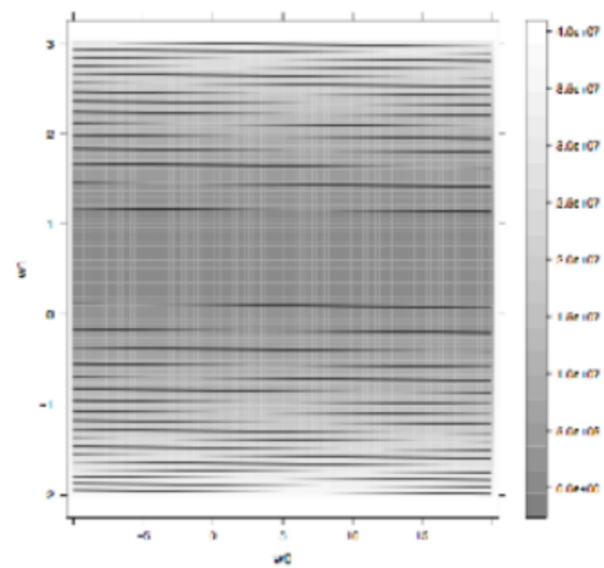
| ID | RENTAL PRICE | Model Prediction | Error | Squared Error |
|---|---|---|---|---|
| 1 | 320 | 316.79 | 3.21 | 10.32 |
| 2 | 380 | 347.82 | 32.18 | 1,035.62 |
| 3 | 400 | 391.26 | 8.74 | 76.32 |
| 4 | 390 | 397.47 | -7.47 | 55.80 |
| 5 | 385 | 419.19 | -34.19 | 1,169.13 |
| 6 | 410 | 440.91 | -30.91 | 955.73 |
| 7 | 480 | 484.36 | -4.36 | 19.01 |
| 8 | 600 | 552.63 | 47.37 | 2,243.90 |
| 9 | 570 | 577.46 | -7.46 | 55.59 |
| 10 | 620 | 627.11 | -7.11 | 50.51 |
| | | | **Sum** | **5,671.64** |
| | | | **Sum of squared errors (Sum/2)** | **2,835.82** |

with $\mathbf{w}[0] = 6.47$ and $\mathbf{w}[1] = 0.62$

For every possible combination of weights, **w**[0] and **w**[1], there is a corresponding sum of squared errors value that can be joined together to make a surface.




(a)



(b)

# Error Surface

The *x*-*y* plane is known as a **weight space** and the surface is known as an **error surface**.

The model that best fits the training data is the model corresponding to the lowest point on the error surface.

### Optimal Condition

$$\frac{\partial}{\partial \mathbf{w}[0]} \frac{1}{2} \sum_{i=1}^{n} (t_i - (\mathbf{w}[0] + \mathbf{w}[1] \times \mathbf{d}_i[1]))^2 = 0$$

and

$$\frac{\partial}{\partial \mathbf{w}[1]} \frac{1}{2} \sum_{i=1}^{n} (t_i - (\mathbf{w}[0] + \mathbf{w}[1] \times \mathbf{d}_i[1]))^2 = 0$$

A **guided search** approach known as the **gradient descent** algorithm for finding optimality

| ID | SIZE | FLOOR | BROADBAND RATE | ENERGY RATING | RENTAL PRICE |
|---|---|---|---|---|---|
| 1 | 500 | 4 | 8 | C | 320 |
| 2 | 550 | 7 | 50 | A | 380 |
| 3 | 620 | 9 | 7 | A | 400 |
| 4 | 630 | 5 | 24 | B | 390 |
| 5 | 665 | 8 | 100 | C | 385 |
| 6 | 700 | 4 | 8 | B | 410 |
| 7 | 770 | 10 | 7 | B | 480 |
| 8 | 880 | 12 | 50 | A | 600 |
| 9 | 920 | 14 | 8 | C | 570 |
| 10 | 1,000 | 9 | 24 | B | 620 |

We can define a multivariate linear regression model as:

$$\begin{aligned}\mathbb{M}_{\mathbf{w}}(\mathbf{d}) &= \mathbf{w}[0] + \mathbf{w}[1] \times \mathbf{d}[1] + \cdots + \mathbf{w}[m] \times \mathbf{d}[m] \\ &= \mathbf{w}[0] + \sum_{j=1}^{m} \mathbf{w}[j] \times \mathbf{d}[j]\end{aligned}$$

Example

$$\begin{aligned}\text{RENTAL PRICE} = \mathbf{w}[0] &+ \mathbf{w}[1] \times \text{SIZE} + \mathbf{w}[2] \times \text{FLOOR} \\ &+ \mathbf{w}[3] \times \text{BROADBAND RATE}\end{aligned}$$

We can make the equation abovelook a little neater by inventing a dummy descriptive feature, **d**[0], that is always equal to 1:

$$\begin{aligned}\mathbb{M}_{\mathbf{w}}(\mathbf{d}) &= \mathbf{w}[0] \times \mathbf{d}[0] + \mathbf{w}[1] \times \mathbf{d}[1] + \ldots \\ &= \sum_{j=0}^{m} \mathbf{w}[j] \times \mathbf{d}[j] \\ &= \mathbf{w} \cdot \mathbf{d}\end{aligned}$$

The sum of squared errors loss function

$$\begin{aligned}L_2(\mathbb{M}_{\mathbf{w}}, \mathcal{D}) &= \frac{1}{2}\sum_{i=1}^{n}(t_i - \mathbb{M}_{\mathbf{w}}(\mathbf{d}_i))^2 \\ &= \frac{1}{2}\sum_{i=1}^{n}(t_i - (\mathbf{w} \cdot \mathbf{d}_i))^2\end{aligned}$$

$$\mathbf{w}[j] \leftarrow \mathbf{w}[j] + \alpha \times \textbf{errorDelta}(\mathcal{D}, \mathbf{w}[j])$$

- Each weight is considered independently and for each one a small adjustment is made by adding a small **delta** value to the current weight, $\mathbf{w}[j]$.
- This adjustment should ensure that the change in the weight leads to a move *downwards* on the error surface.

Imagine for a moment that our training dataset, D contains just one training example: ($\mathbf{d}$, $t$ )

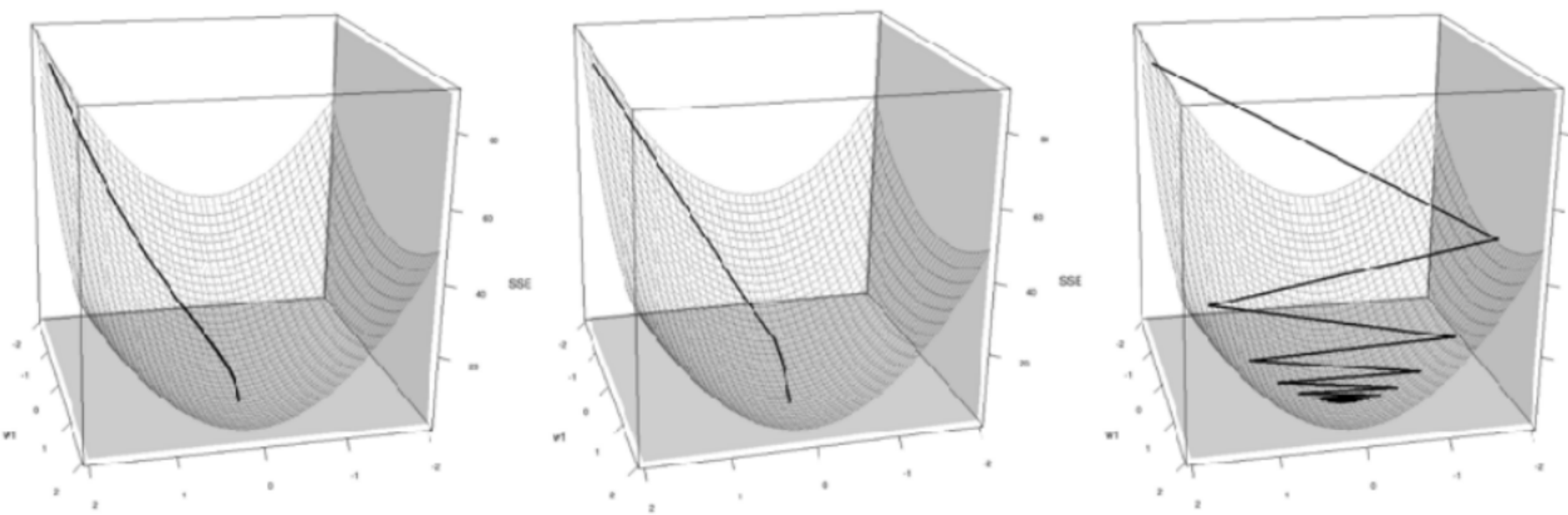
$$
\begin{aligned}
\frac{\partial}{\partial \mathbf{w}[j]} L_2(\mathbb{M}_{\mathbf{w}}, \mathcal{D}) &= \frac{\partial}{\partial \mathbf{w}[j]} \left( \frac{1}{2} (t - \mathbb{M}_{\mathbf{w}}(\mathbf{d}))^2 \right) \\
&= (t - \mathbb{M}_{\mathbf{w}}(\mathbf{d})) \times \frac{\partial}{\partial \mathbf{w}[j]} (t - \mathbb{M}_{\mathbf{w}}(\mathbf{d})) \\
&= (t - \mathbb{M}_{\mathbf{w}}(\mathbf{d})) \times \frac{\partial}{\partial \mathbf{w}[j]} (t - (\mathbf{w} \cdot \mathbf{d})) \\
&= (t - \mathbb{M}_{\mathbf{w}}(\mathbf{d})) \times -\mathbf{d}[j]
\end{aligned}
$$

multiple training instances

$$
-\frac{\partial}{\partial \mathbf{w}[j]} L_2(\mathbb{M}_{\mathbf{w}}, \mathcal{D}) = \sum_{i=1}^{n} ((t_i - \mathbb{M}_{\mathbf{w}}(\mathbf{d}_i)) \times \mathbf{d}_i[j])
$$

$$\mathbf{w}[j] \leftarrow \mathbf{w}[j] + \alpha \underbrace{\sum_{i=1}^{n} ((t_i - \mathbb{M}_{\mathbf{w}}(\mathbf{d}_i)) \times d_i[j])}_{errorDelta(\mathcal{D}, \mathbf{w}[j])}$$

The learning rate α determines the size of the adjustment made to each weight at each step in the process.



結論：

- 把問題描述程方程式
- 可以求出cost function

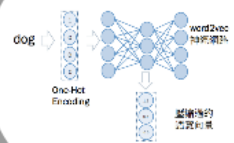
# 實作

機器學習的原理

跟小學生考數學一樣

2X+Y=10

Word2 vec




## Parameterized Prediction model

A **parameterized** prediction model is initialized with a set of random parameters and an error function is used to judge how well this initial model performs when making predictions for instances in a training dataset.

Based on the value of the error function the parameters are iteratively adjusted to create a more and more accurate model.

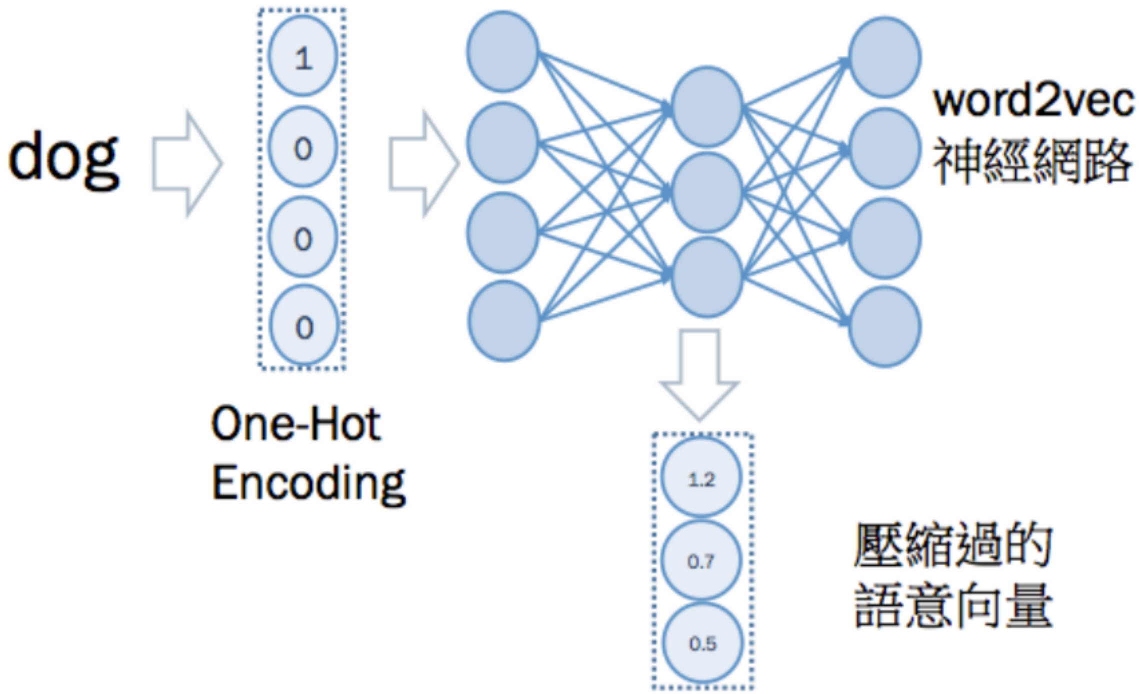
$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

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

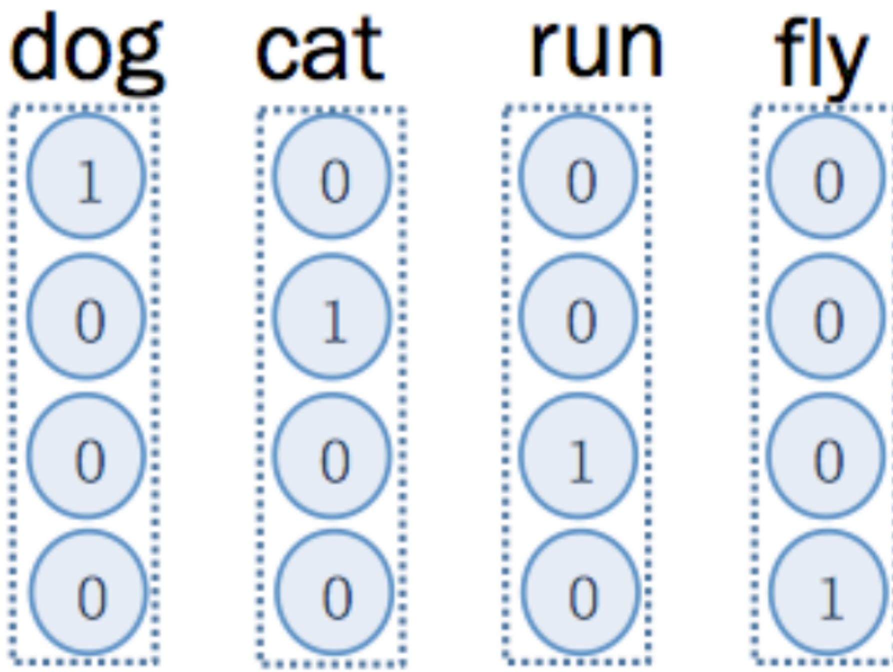
**Cost Function based L2 Norm**

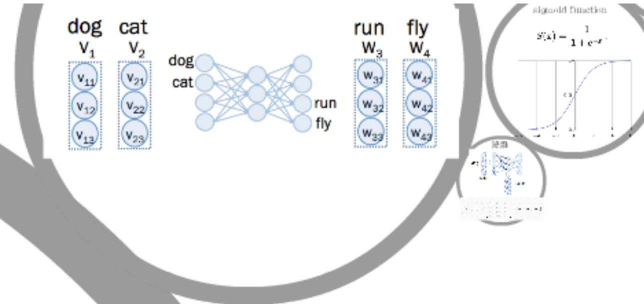
| ID | Size | Floor | Broadband Rate | Energy Rating | Rental Price |
|---|---|---|---|---|---|
| 1 | 500 | 4 | 8 | C | 320 |
| 2 | 550 | 7 | 50 | A | 380 |
| 3 | 620 | 9 | 7 | A | 400 |
| 4 | 630 | 5 | 24 | B | 390 |
| 5 | 665 | 8 | 100 | C | 385 |
| 6 | 700 | 4 | 8 | B | 410 |
| 7 | 770 | 10 | 7 | B | 480 |
| 8 | 880 | 12 | 50 | A | 600 |
| 9 | 920 | 14 | 8 | C | 570 |

# Word2 vec


dog
1
0
0
0
One-Hot
Encoding
word2vec
神經網路
1.2
0.7
0.5
壓縮過的
語意向量

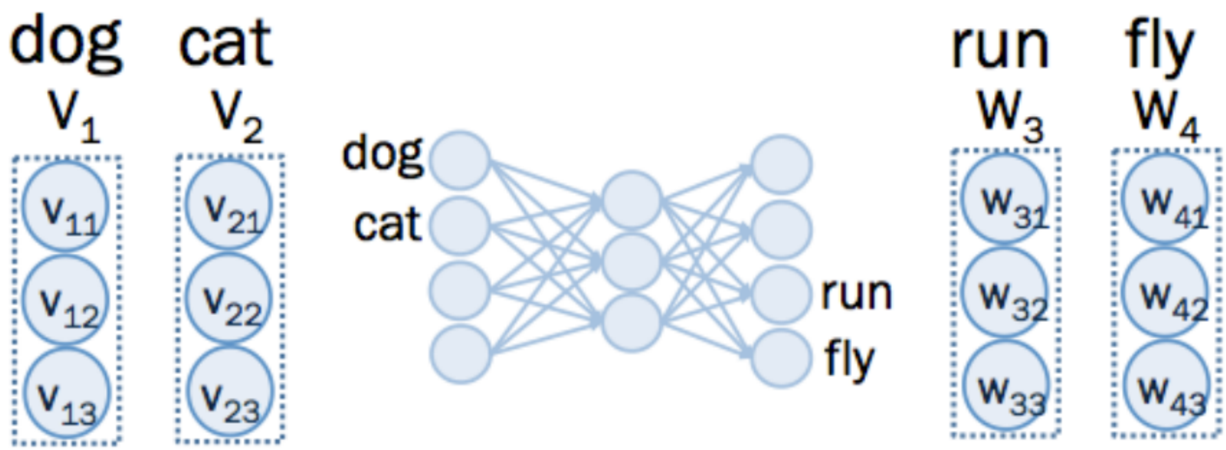

facebook
Scrapy
字幕
句子也可以轉成向量
叫做sentence2vec
dog cat run fly

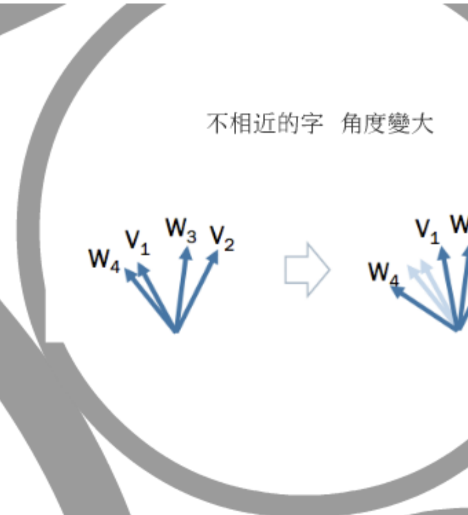


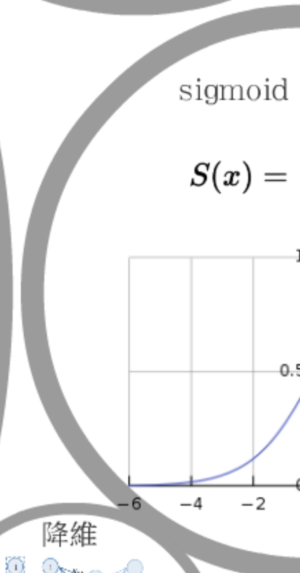



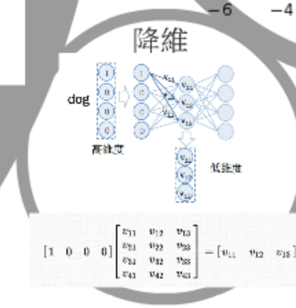

Prezi

# 字詞前後關係

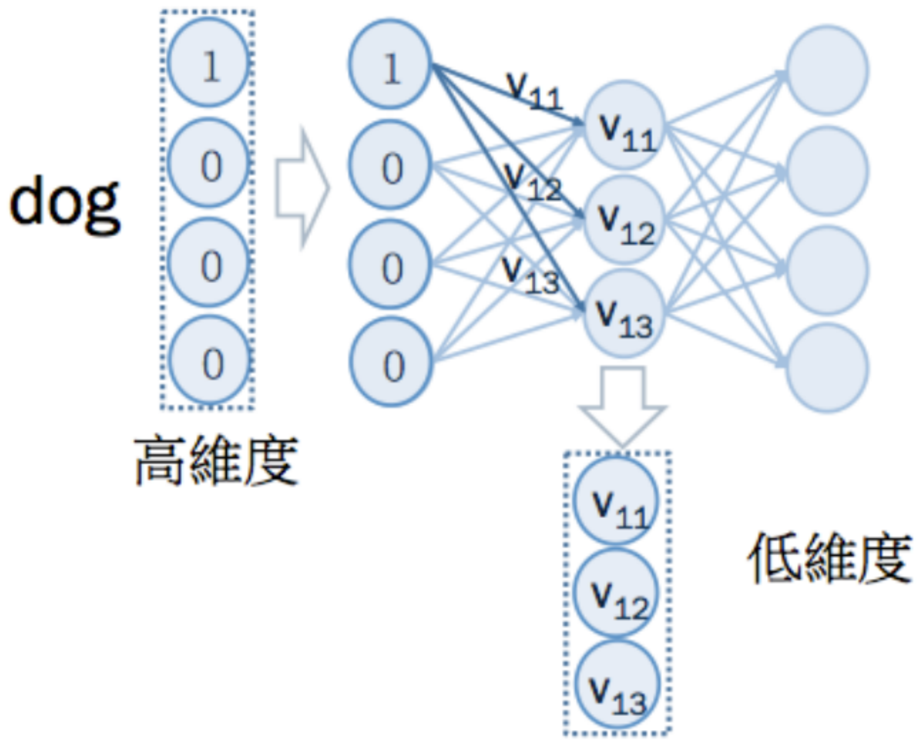

dog
cat
V1
V2
v11
v12
v13
v21
v22
v23
dog
cat
run
fly
run
fly
W3
W4
w31
w32
w33
w41
w42
w43


不相近的字 角度變大
W4
V1
W3
V2


sigmoid
S(x) =


降維
dog

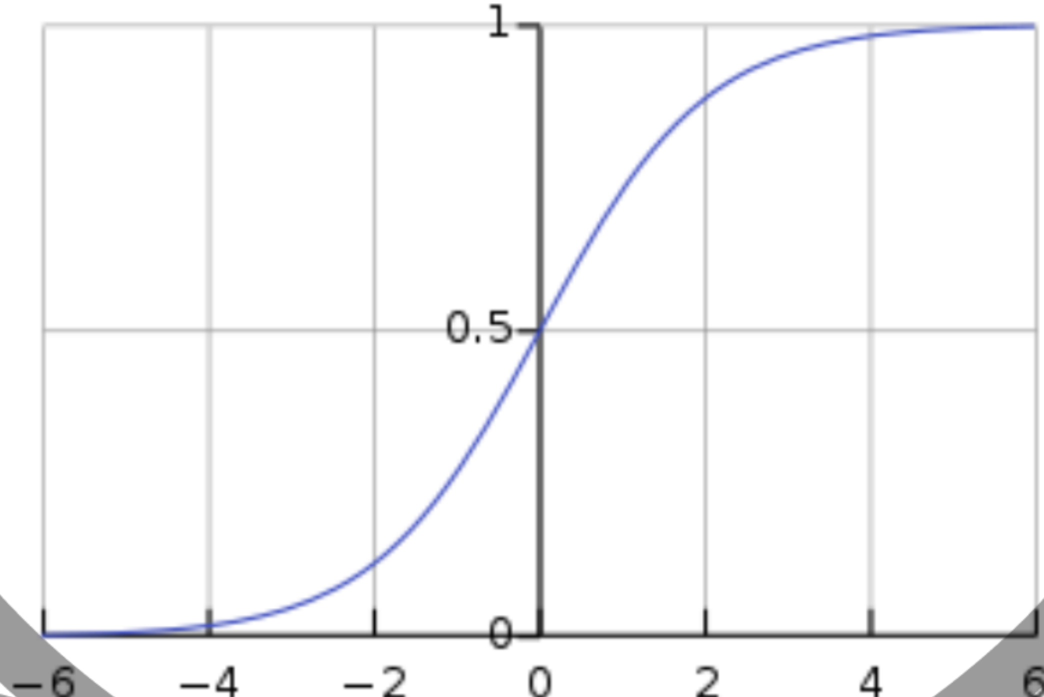
# 降維



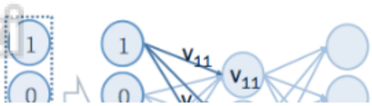

$$\begin{bmatrix}1 & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}v_{11} & v_{12} & v_{13}\\ v_{21} & v_{22} & v_{23}\\ v_{31} & v_{32} & v_{33}\\ v_{41} & v_{42} & v_{43}\end{bmatrix} = \begin{bmatrix}v_{11} & v_{12} & v_{13}\end{bmatrix}$$

sigmoid function

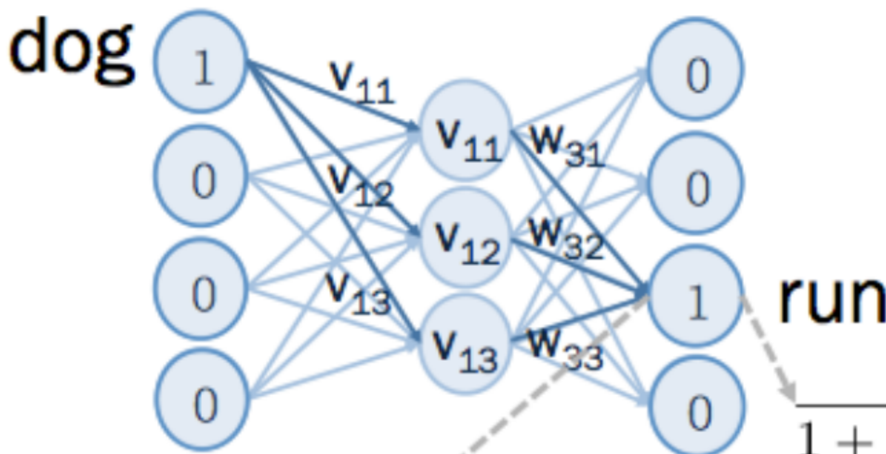
$$S(x) = \frac{1}{1+e^{-x}}.$$




降維



$V_2$

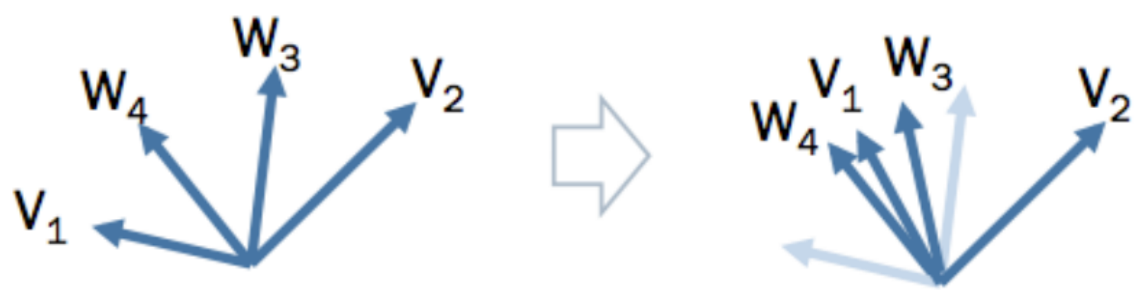


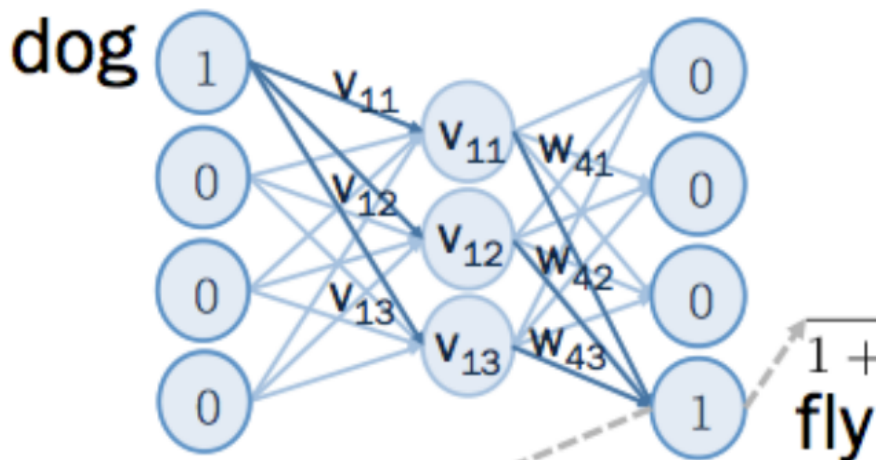

$$\frac{1}{1+e^{-\mathbf{V}_1\mathbf{W}_3}} \approx 1$$

$$\mathbf{V}_1 \cdot \mathbf{W}_3 = v_{11}w_{31} + v_{12}w_{32} + v_{13}w_{33}$$

unction

# 相似的字 夾角變小






$$\frac{1}{1+e^{-\mathbf{V}_1\mathbf{W}_4}} \approx 0$$

$$\mathbf{V}_1 \cdot \mathbf{W}_4 = v_{11}w_{41} + v_{12}w_{42} + v_{13}w_{43}$$

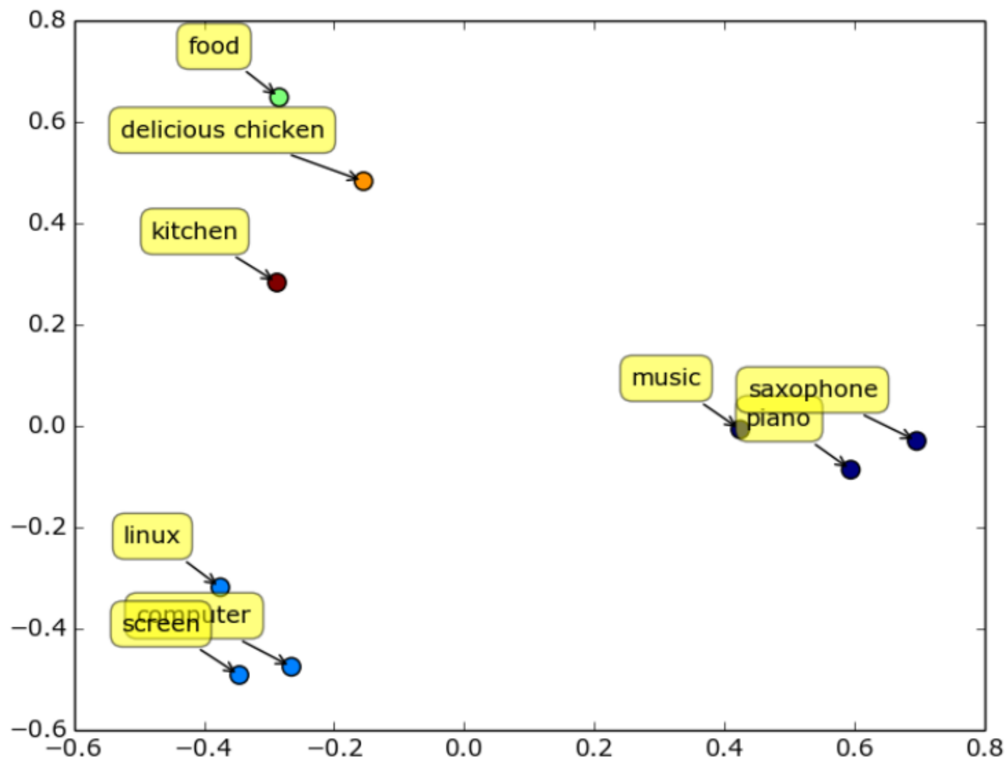
不相近的字 角度變大


相似的字 夾角變小

dog

$V_1 \cdot W$

# 結果:



food
delicious chicken
kitchen
music
saxophone
piano
linux
computer
screen
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8

句子也可以轉成向量

叫做sentence2vec

Scrapy

例如:


句子也可以轉成向量

sentence2vec

```
1
00:00:00,500 --> 00:00:04,430
平匡　你可以成為我的男朋友嗎

2
00:00:04,430 --> 00:00:05,990
我和你成為戀人嗎

3
00:00:05,990 --> 00:00:07,620
和你交往的話

4
00:00:07,620 --> 00:00:11,460
既不用對周圍隱瞞　也不用刻意注意些什麼
```

# 粉專

django

# 使用到的技術

python™

mongoDB®

使用到的套件

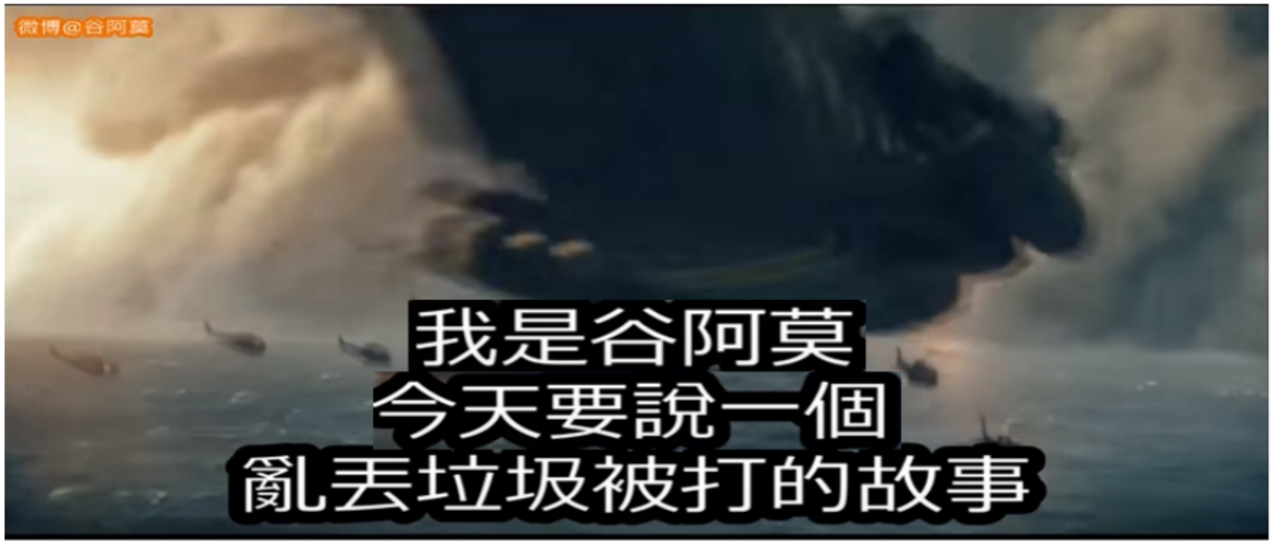
sentence2vec

Scrapy

最後我們的女友機器人算是失敗了

# 失敗的經驗分享


PTT留言沒有水準



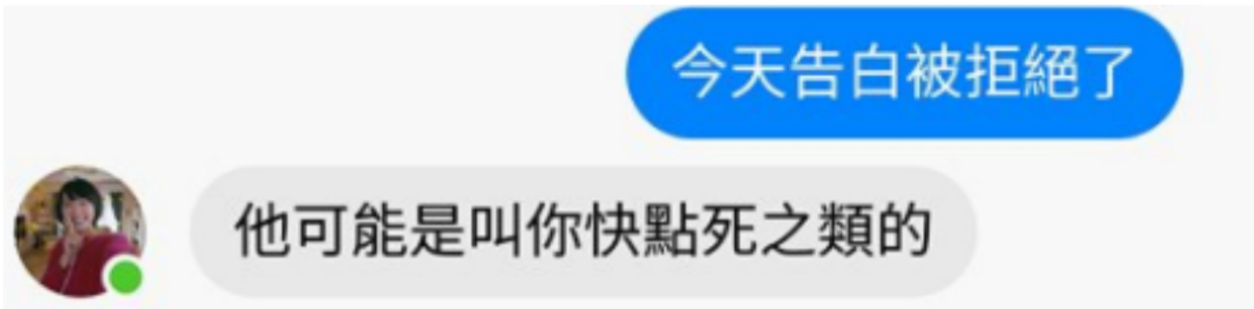

電影內容太廣泛


訓練的過程沒有很好的
benchmark (指標)來判斷好壞

# 還有


不要把作業拖到最後一天...

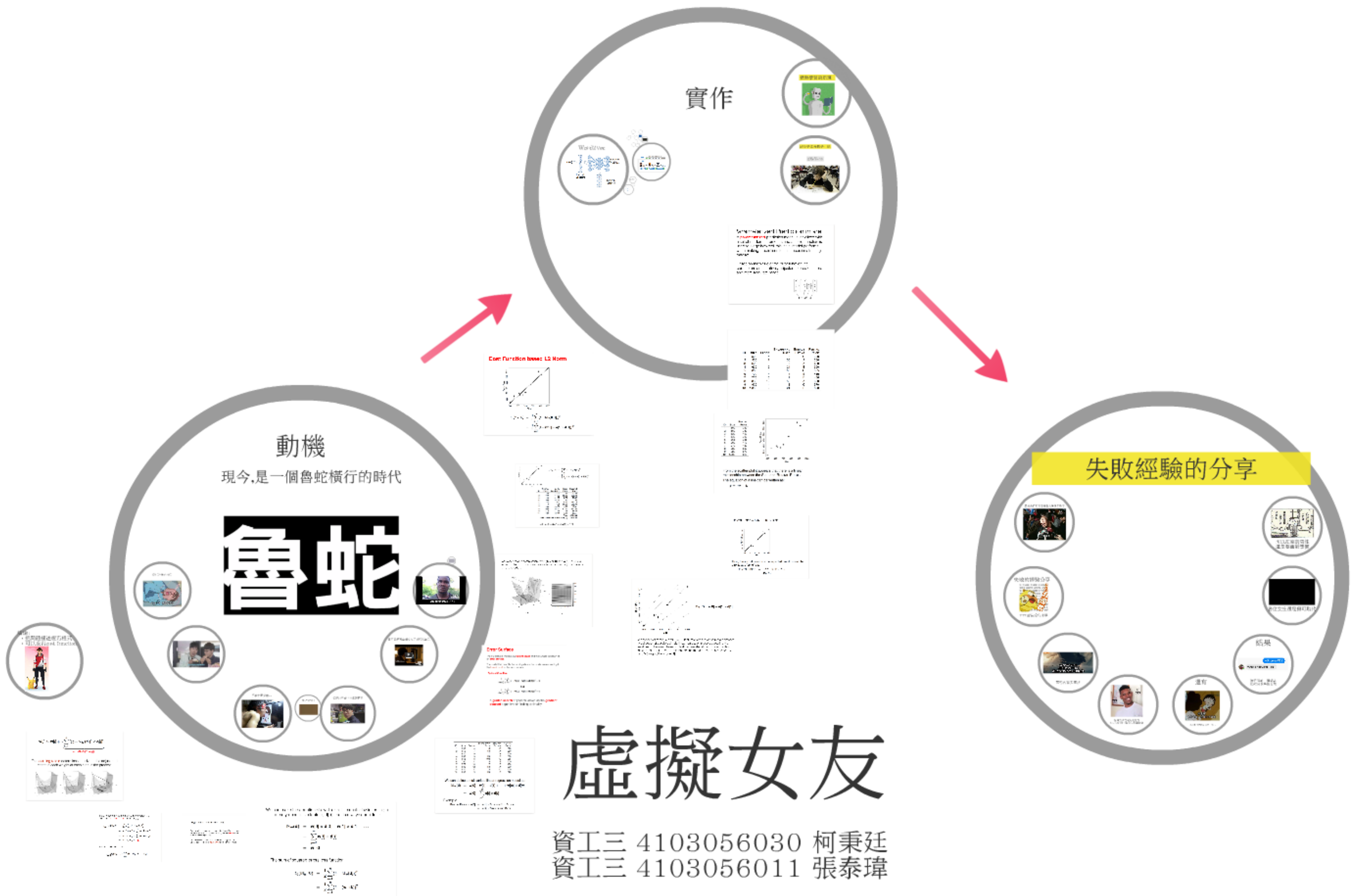
# 結果




我們得到一個講話
比較沒水準的女友


各位女生還是無可取代


所以在座的男性
還是要面對事實


實作
動機
現今,是一個魯蛇橫行的時代
魯蛇
失敗經驗的分享
虛擬女友
資工三 4103056030 柯秉廷
資工三 4103056011 張泰瑋
Cost Function based L2 Norm