Classification: Probabilistic Generative Model

Classification

x ➡ **Function** ➡ Class n

- Credit Scoring
  - Input: income, savings, profession, age, past financial history ……
  - Output: accept or refuse

- Medical Diagnosis
  - Input: current symptoms, age, gender, past medical history ……
  - Output: which kind of diseases

- Handwritten character recognition

Input: 　　output: 金

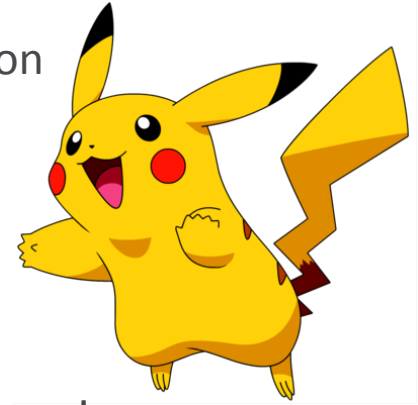- Face recognition
  - Input: image of a face, output: person

Example Application

POKÉMON TYPE SYMBOLS

NORMAL FIRE WATER ELECTRIC GRASS ICE

FIGHTING POISON GROUND FLYING PSYCHIC BUG

ROCK GHOST DRAGON DARK STEEL FAIRY

$f(\ \ ) = \ \ \ \ f(\ \ ) = \ \ \ \ f(\ \ ) = $

Example Application

- **Total**: sum of all stats that come after this, a general guide to how strong a pokemon is   320

- **HP**: hit points, or health, defines how much damage a pokemon can withstand before fainting   35

- **Attack**: the base modifier for normal attacks (eg. Scratch, Punch)   55

- **Defense**: the base damage resistance against normal attacks   40

- **SP Atk**: special attack, the base modifier for special attacks (e.g. fire blast, bubble beam)   50

- **SP Def**: the base damage resistance against special attacks   50

- **Speed**: determines which pokemon attacks first each round   90

Can we predict the "type" of pokemon based on the information?

Example Application

×	一般	格門	飛行	毒	地面	岩石	蟲	幽靈	鋼	火	水	草	電	超能力	冰	龍	惡	妖精
一般	1×	1×	1×	1×	1×	½×	1×	0×	½×	1×	1×	1×	1×	1×	1×	1×	1×	1×
格門	2×	1×	½×	½×	1×	2×	½×	0×	2×	1×	1×	1×	1×	½×	2×	1×	2×	½×
飛行	1×	2×	1×	1×	1×	½×	2×	1×	½×	1×	1×	2×	½×	1×	1×	1×	1×	1×
毒	1×	1×	1×	½×	½×	½×	1×	½×	0×	1×	1×	2×	1×	1×	1×	1×	1×	2×
地面	1×	1×	0×	2×	1×	2×	½×	1×	2×	2×	1×	½×	2×	1×	1×	1×	1×	1×
岩石	1×	½×	2×	1×	½×	1×	2×	1×	½×	2×	1×	1×	1×	2×	1×	1×	1×	1×
蟲	1×	½×	½×	½×	1×	1×	1×	½×	½×	½×	1×	2×	1×	2×	1×	1×	2×	½×
幽靈	0×	1×	1×	1×	1×	1×	1×	2×	1×	1×	1×	1×	1×	2×	1×	1×	½×	1×
鋼	1×	1×	1×	1×	1×	2×	1×	1×	½×	½×	½×	1×	½×	1×	2×	1×	1×	2×
火	1×	1×	1×	1×	1×	½×	2×	1×	2×	½×	½×	2×	1×	1×	2×	½×	1×	1×
水	1×	1×	1×	1×	2×	2×	1×	1×	1×	2×	½×	½×	1×	1×	1×	½×	1×	1×
草	1×	1×	½×	½×	2×	2×	½×	1×	½×	½×	2×	½×	1×	1×	1×	½×	1×	1×
電	1×	1×	2×	1×	0×	1×	1×	1×	1×	1×	2×	½×	½×	1×	1×	½×	1×	1×
超能力	1×	2×	1×	2×	1×	1×	1×	1×	½×	1×	1×	1×	1×	½×	1×	1×	0×	1×
冰	1×	1×	2×	1×	2×	1×	1×	1×	½×	½×	½×	2×	1×	1×	½×	2×	1×	1×
龍	1×	1×	1×	1×	1×	1×	1×	1×	½×	1×	1×	1×	1×	1×	1×	2×	1×	0×
惡	1×	½×	1×	1×	1×	1×	1×	2×	1×	1×	1×	1×	1×	2×	1×	1×	½×	½×
妖精	1×	2×	1×	½×	1×	1×	1×	1×	½×	½×	1×	1×	1×	1×	1×	2×	2×	1×

防禦方的屬性

攻擊方的屬性

這些倍數適用於XY及之後的遊戲。

How to do Classification

- Training data for Classification

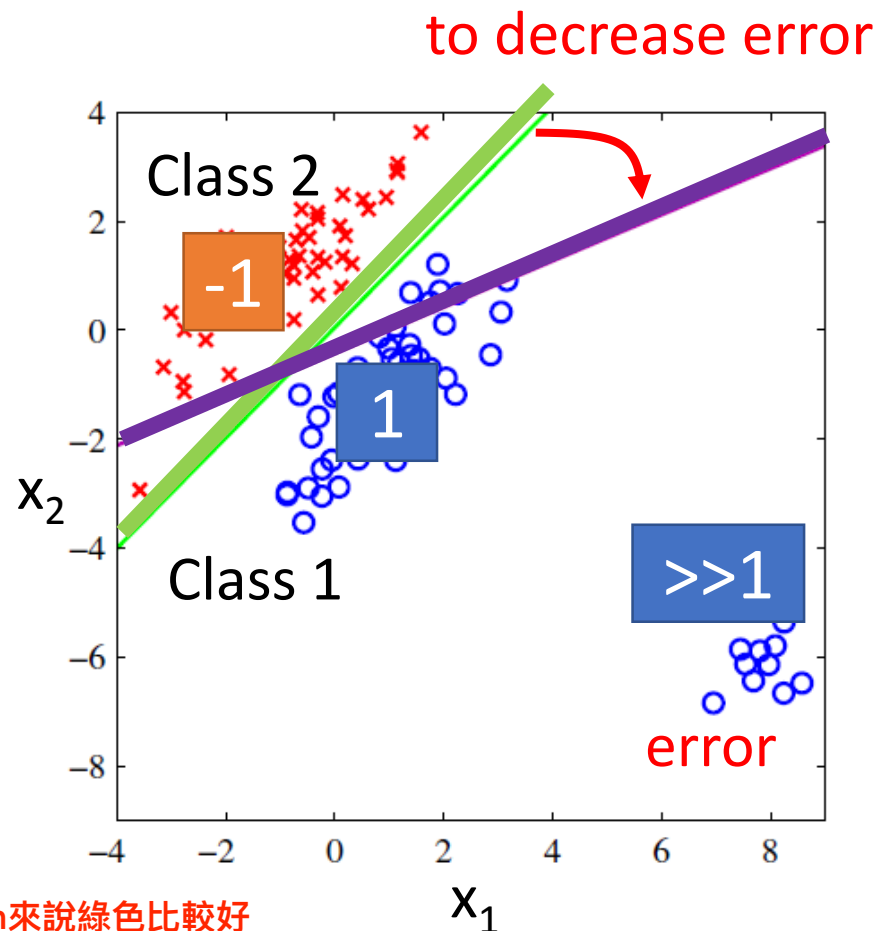$$(x^1, \hat{y}^1) \qquad (x^2, \hat{y}^2) \qquad \cdots\cdots \qquad (x^N, \hat{y}^N)$$
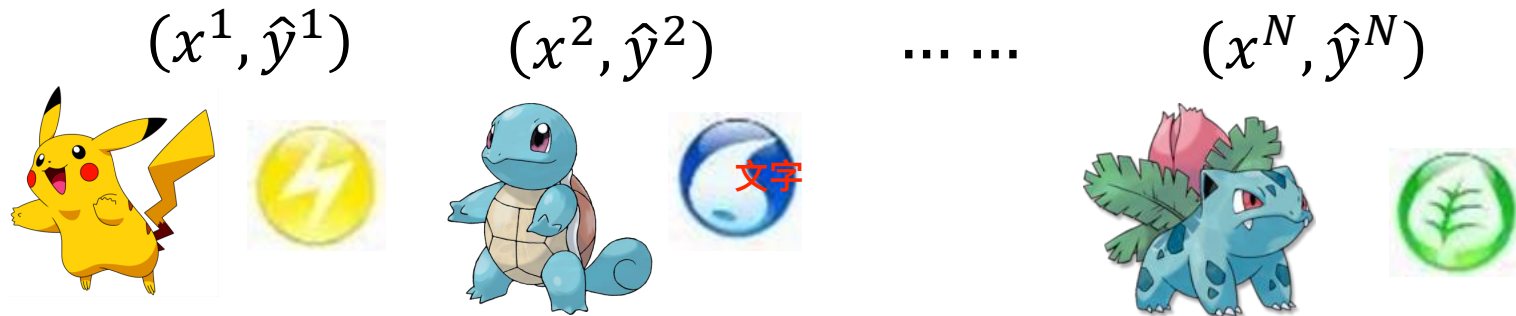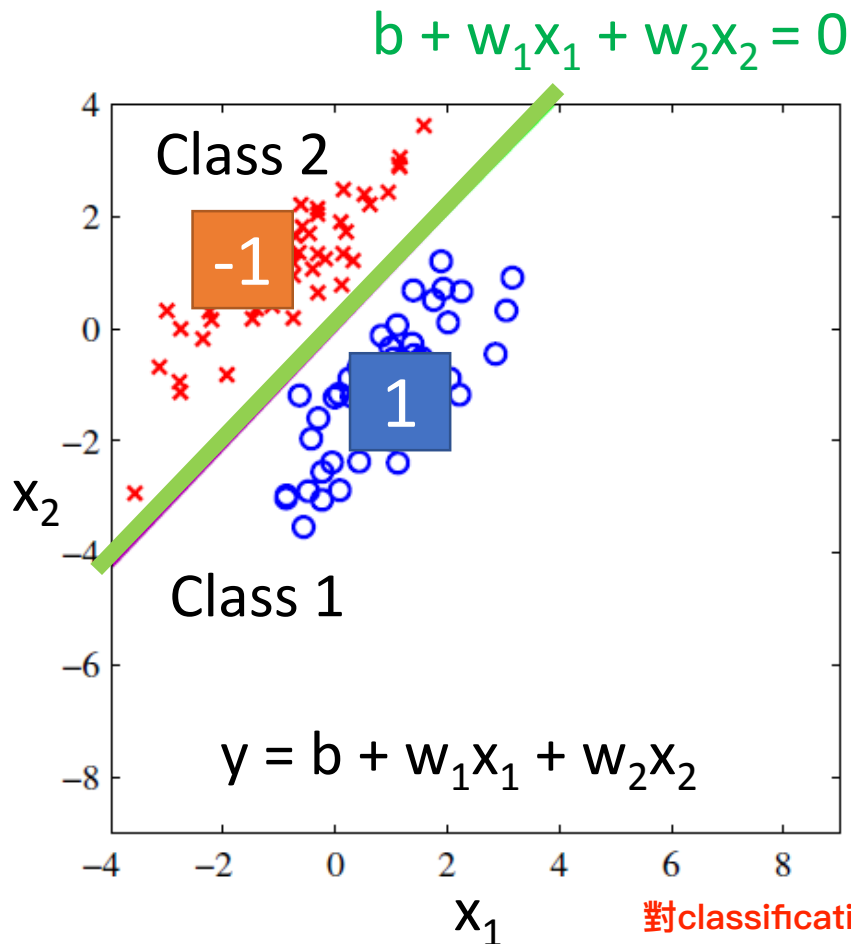

Classification as Regression?

Binary classification as example

Training: Class 1 means the target is 1; Class 2 means the target is -1

Testing: closer to 1 → class 1; closer to -1 → class 2

$b + w_1x_1 + w_2x_2 = 0$

to decrease error

Class 2

-1

1

Class 1

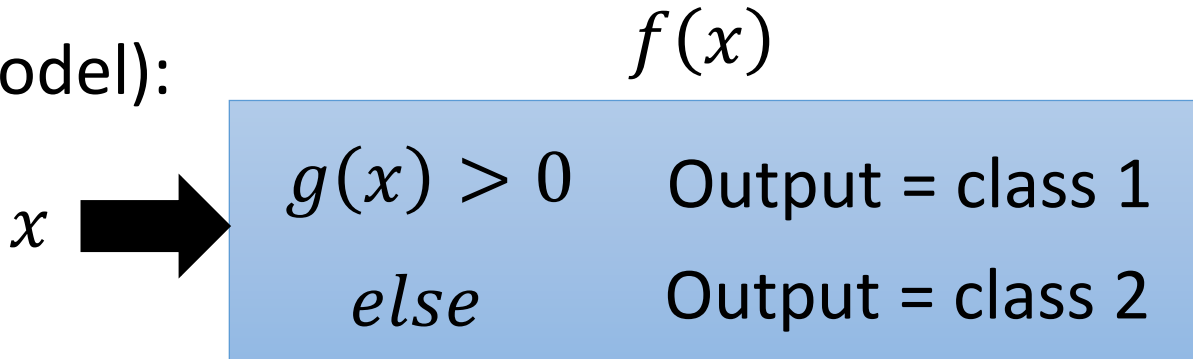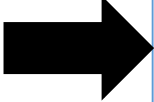$y = b + w_1x_1 + w_2x_2$

Class 2

-1

1

Class 1

>>1

error

對classification來說綠色比較好
但如果利用regularization會修正成紫色，不一定比較好

Penalize to the examples that are "too correct" … (Bishop, P186)

- Multiple class: Class 1 means the target is 1; Class 2 means the target is 2; Class 3 means the target is 3 …… problematic

Ideal Alternatives

- Function (Model):

$$f(x)$$

$$x \rightarrow \boxed{\begin{array}{ll} g(x) > 0 & \text{Output = class 1} \\ else & \text{Output = class 2} \end{array}}$$

- Loss function:

$$L(f) = \sum_{n} \delta(f(x^n) \neq \hat{y}^n)$$

若不相等則output 1
相等則output為0
->統計分類錯誤次數

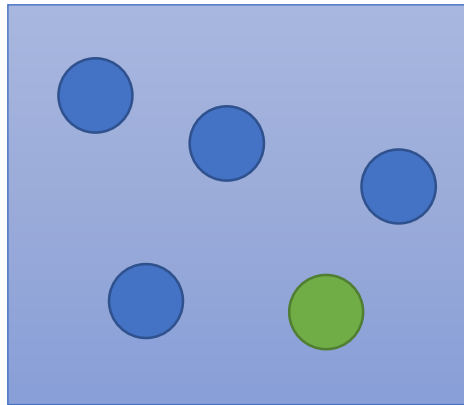The number of times f get incorrect results on training data.

- Find the best function:
  - Example: Perceptron, SVM

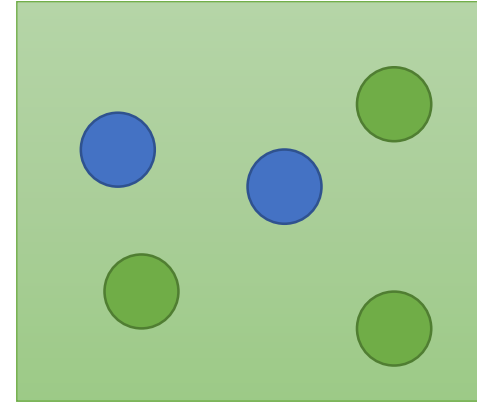Not Today

Two Boxes

Box 1

$P(B_1) = 2/3$

$P(\text{Blue}|B_1) = 4/5$
$P(\text{Green}|B_1) = 1/5$

Box 2

$P(B_2) = 1/3$

$P(\text{Blue}|B_1) = 2/5$
$P(\text{Green}|B_1) = 3/5$

from one of the boxes

Where does it come from?

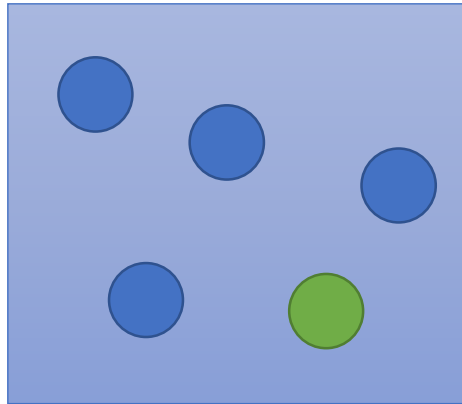$$P(B_1 \mid \text{Blue}) = \frac{P(\text{Blue}|B_1)P(B_1)}{P(\text{Blue}|B_1)P(B_1) + P(\text{Blue}|B_2)P(B_2)}$$

Two Classes

Class 1

$P(C_1)$

$P(x|C_1)$

Class 2

$P(C_2)$
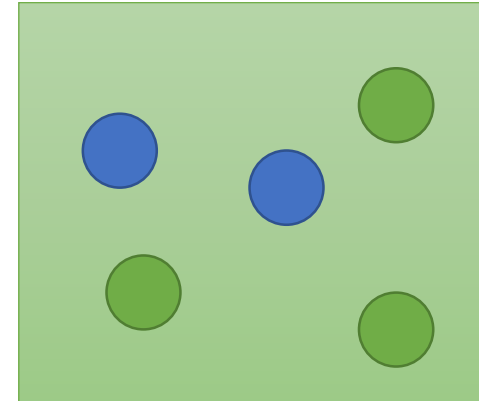
$P(x|C_2)$

Given an x, which class does it belong to

$$P(C_1|x) = \frac{P(x|C_1)P(C_1)}{P(x|C_1)P(C_1) + P(x|C_2)P(C_2)}$$

Generative Model $\quad P(x) = P(x|C_1)P(C_1) + P(x|C_2)P(C_2)$

Prior

Class 1

$P(C_1)$

Water

Class 2

$P(C_2)$
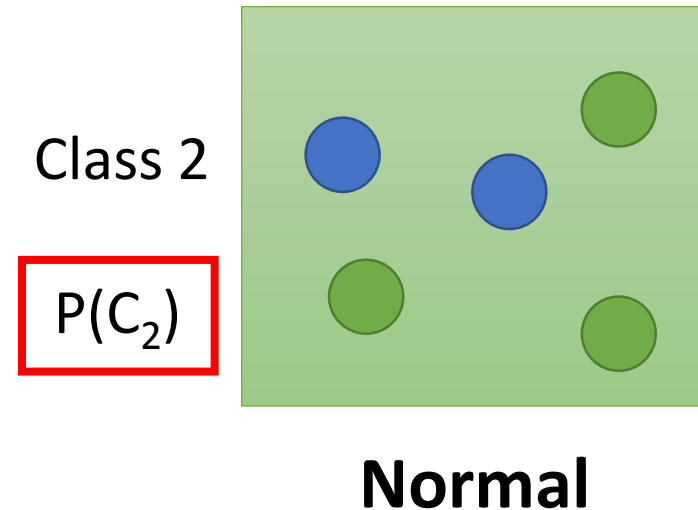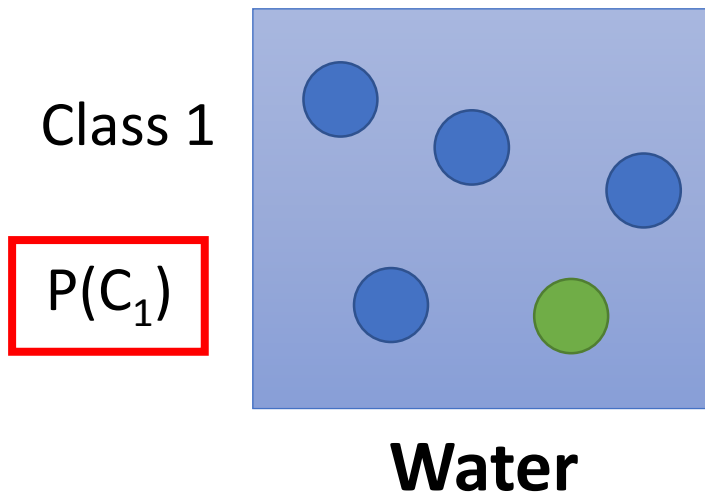
Normal

Water and Normal type with ID < 400 for training, rest for testing
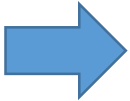
Training: 79 Water, 61 Normal

$P(C_1) = 79 / (79 + 61) = 0.56$

$P(C_2) = 61 / (79 + 61) = 0.44$

Probability from Class

$P(x|C_1) = ?$ $P($ $|Water) = ?$

Each Pokémon is represented as a <u>vector</u> by its attribute.   ➡   feature

Water Type   79 in total

Probability from Class - Feature

- Considering **Defense** and **SP Defense**
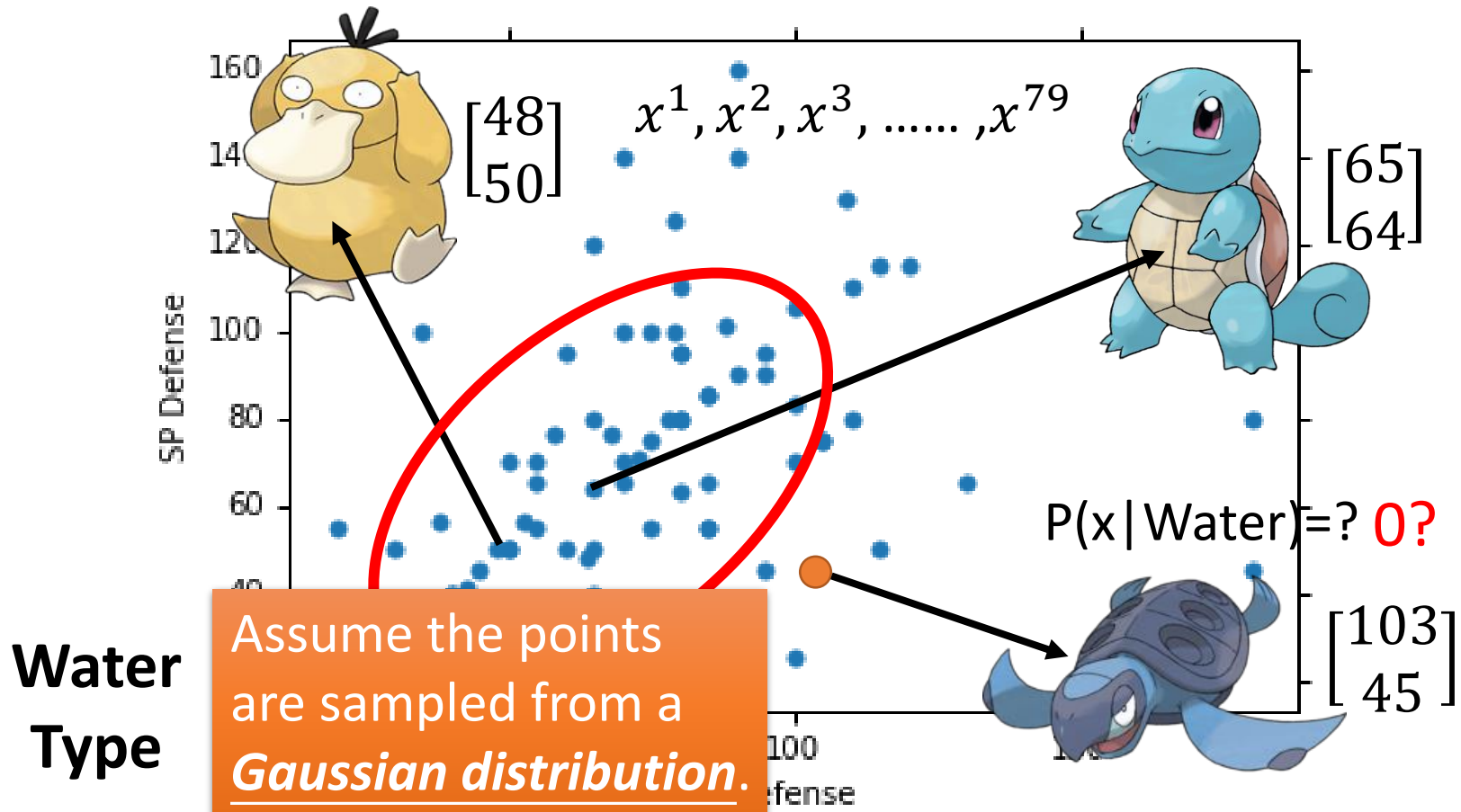
$\begin{bmatrix} 48 \\ 50 \end{bmatrix}$ $x^1, x^2, x^3, \ldots\ldots, x^{79}$

$\begin{bmatrix} 65 \\ 64 \end{bmatrix}$

P(x|Water)=? 0?

$\begin{bmatrix} 103 \\ 45 \end{bmatrix}$

Water Type

Assume the points are sampled from a *Gaussian distribution*.

Gaussian Distribution

$$f_{\mu,\Sigma}(x) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}$$

Input: vector x, output: probability of sampling x

The shape of the function determines by **mean μ** and **covariance matrix Σ**

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

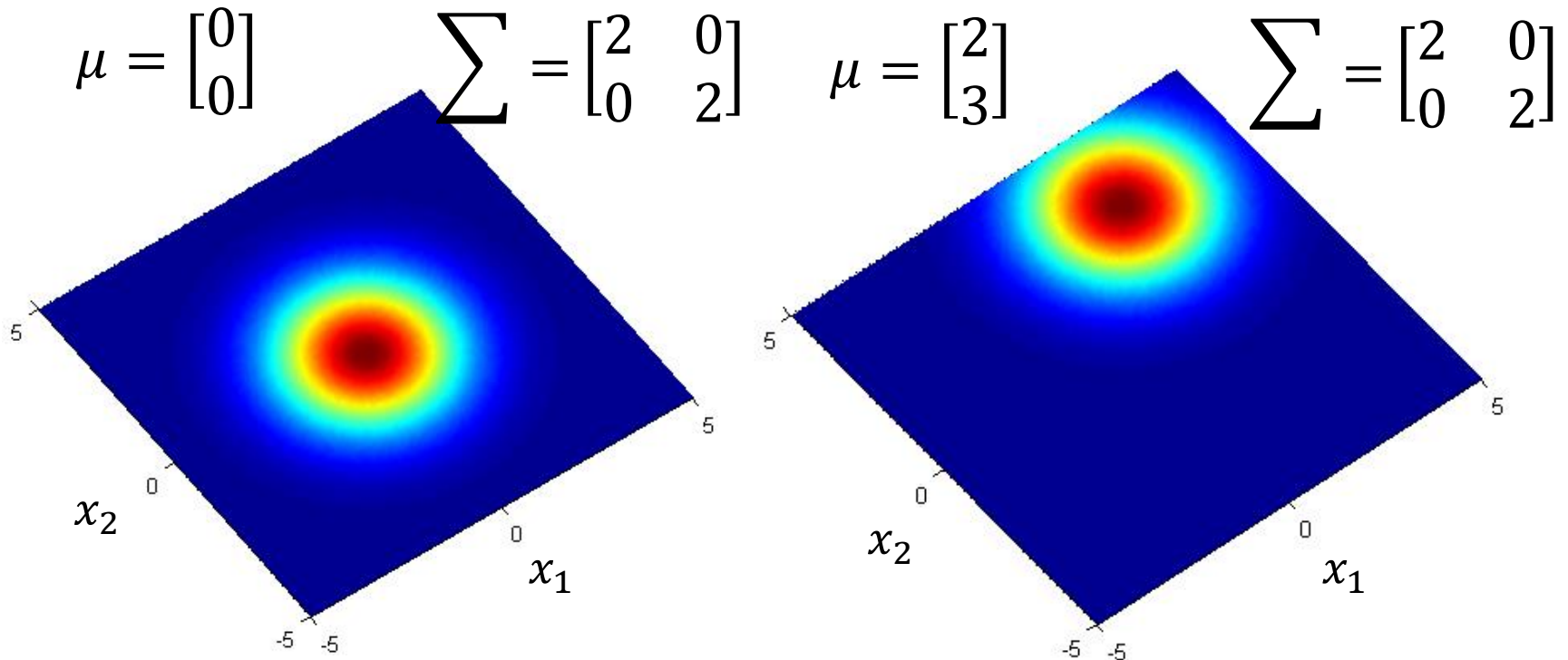$$\mu = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Gaussian Distribution

$$f_{\mu,\Sigma}(x) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$$

Input: vector x, output: probability of sampling x

The shape of the function determines by **mean μ** and **covariance matrix Σ**
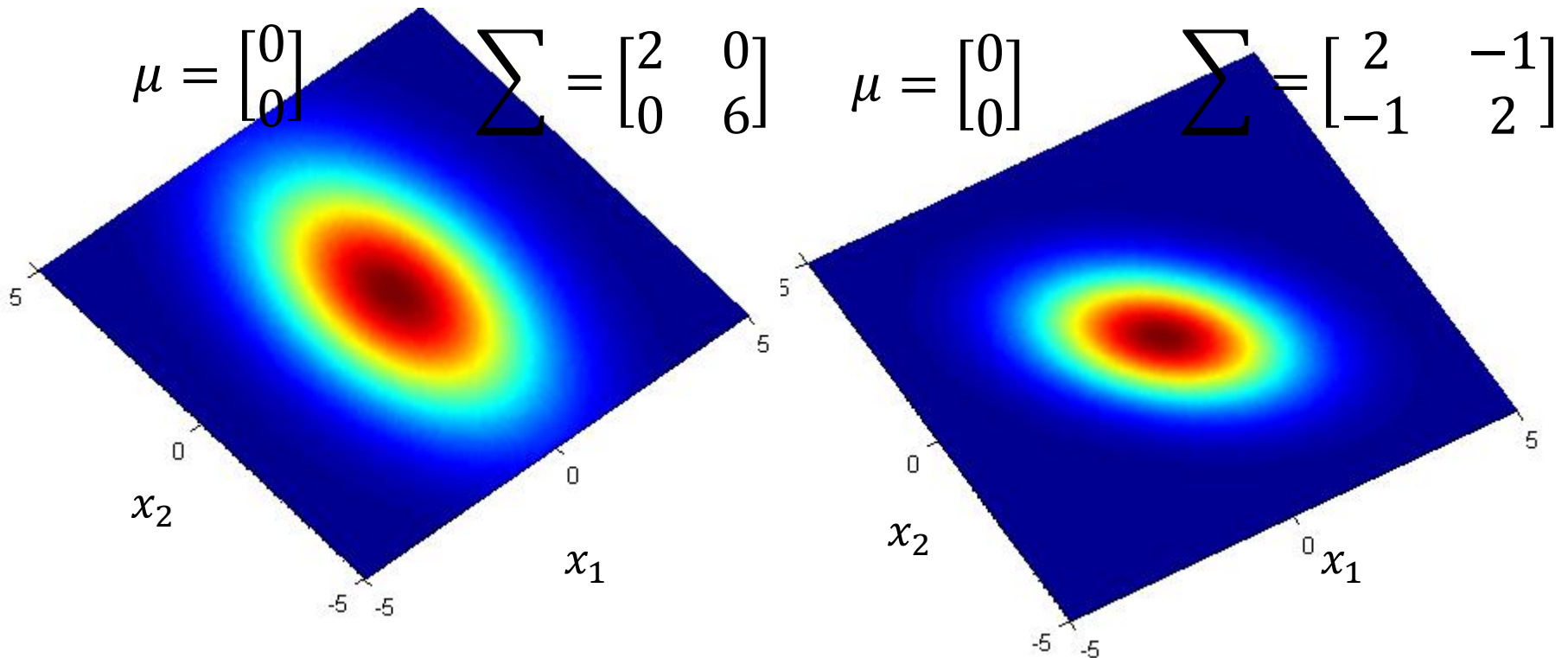
$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ $\sum = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}$ $\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ $\sum = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$

Probability from Class

Assume the points are sampled from a Gaussian distribution

Find the Gaussian distribution behind them ➡ Probability for new points

$x^1, x^2, x^3, \ldots\ldots, x^{79}$

New x

$$\mu = \begin{bmatrix} 75.0 \\ 71.3 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 874 & 327 \\ 327 & 929 \end{bmatrix}$$

How to find them?

Water Type

SP Defense

Defense

$$f_{\mu,\Sigma}(x) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}$$
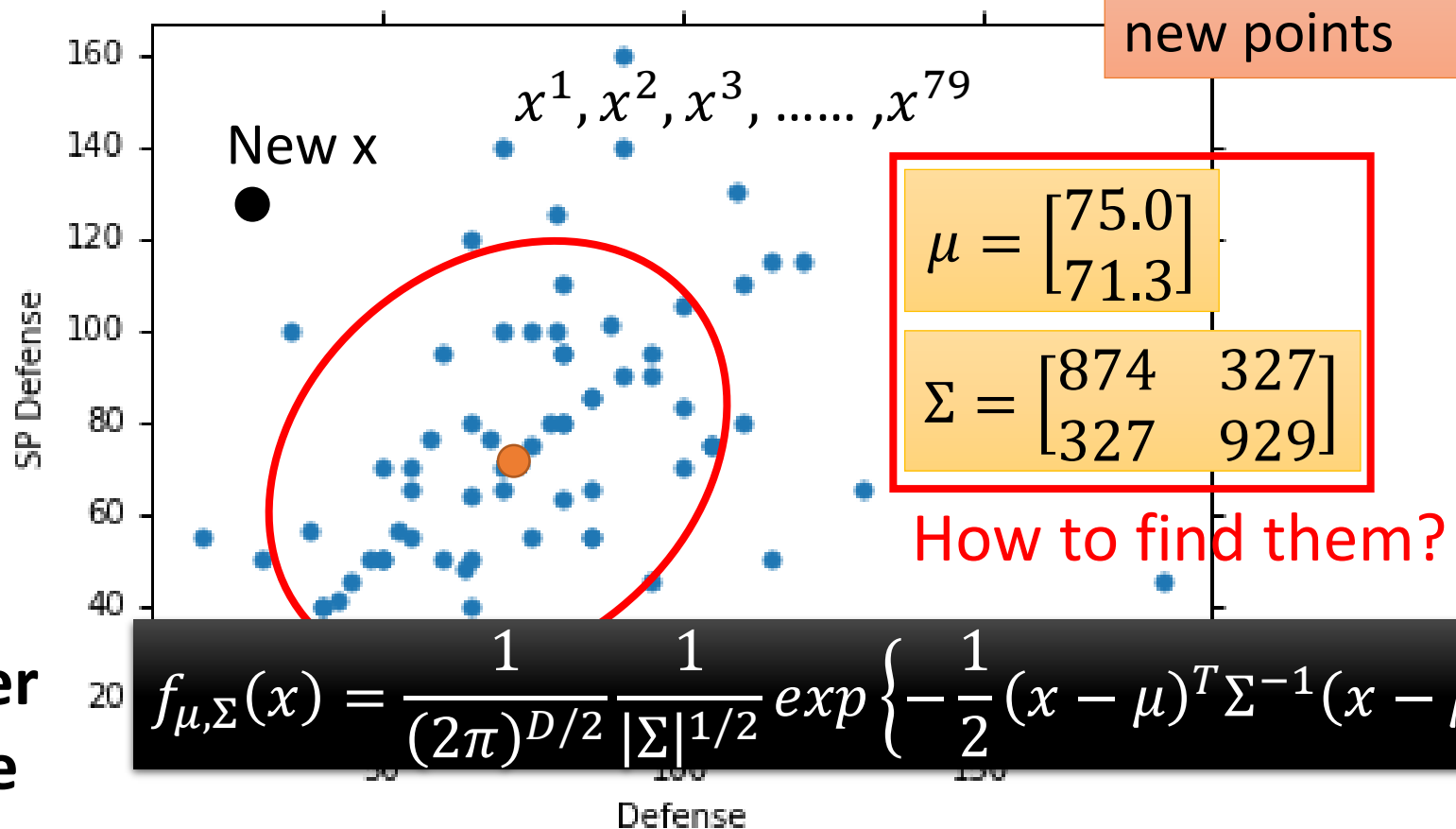
Maximum Likelihood

$$f_{\mu,\Sigma}(x) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}$$

$$\Sigma = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$$

$$x^1, x^2, x^3, \ldots\ldots, x^{79}$$

$$\mu = \begin{bmatrix} 150 \\ 140 \end{bmatrix}$$

$$\mu = \begin{bmatrix} 75 \\ 75 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$
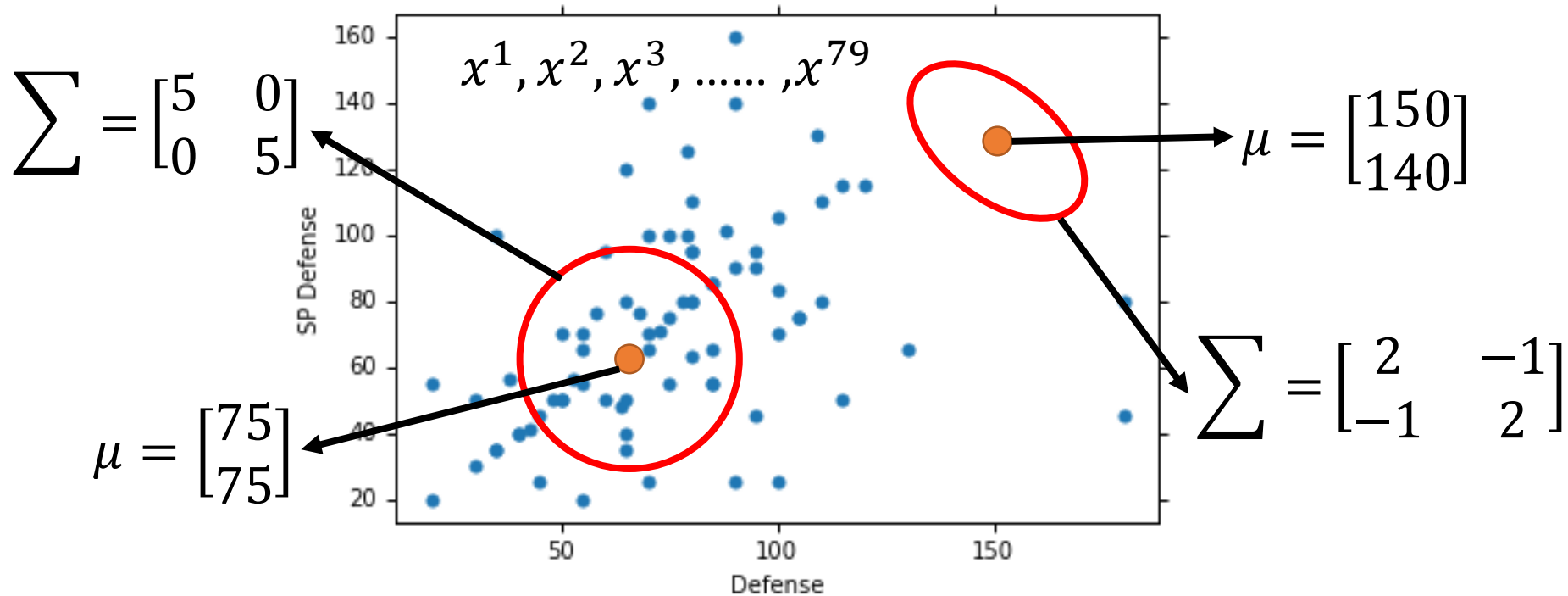
The Gaussian with any mean μ and covariance matrix Σ can generate these points. ⟹ Different Likelihood

Likelihood of a Gaussian with mean μ and covariance matrix Σ

= the probability of the Gaussian samples $x^1, x^2, x^3, \ldots\ldots, x^{79}$

$$L(\mu, \Sigma) = f_{\mu,\Sigma}(x^1)f_{\mu,\Sigma}(x^2)f_{\mu,\Sigma}(x^3)\ldots\ldots f_{\mu,\Sigma}(x^{79})$$

Likelihood

Maximum Likelihood

We have the "Water" type Pokémons: $x^1, x^2, x^3, \ldots\ldots, x^{79}$

We assume $x^1, x^2, x^3, \ldots\ldots, x^{79}$ generate from the Gaussian (μ^*, Σ^*) with the ***maximum likelihood***

找一個Gaussian使得sample出79個點的likelihood最大

$$L(\mu, \Sigma) = f_{\mu,\Sigma}(x^1) f_{\mu,\Sigma}(x^2) f_{\mu,\Sigma}(x^3) \ldots\ldots f_{\mu,\Sigma}(x^{79})$$

likelihood function

$$f_{\mu,\Sigma}(x) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}$$

$$\mu^*, \Sigma^* = arg \max_{\mu,\Sigma} L(\mu, \Sigma)$$

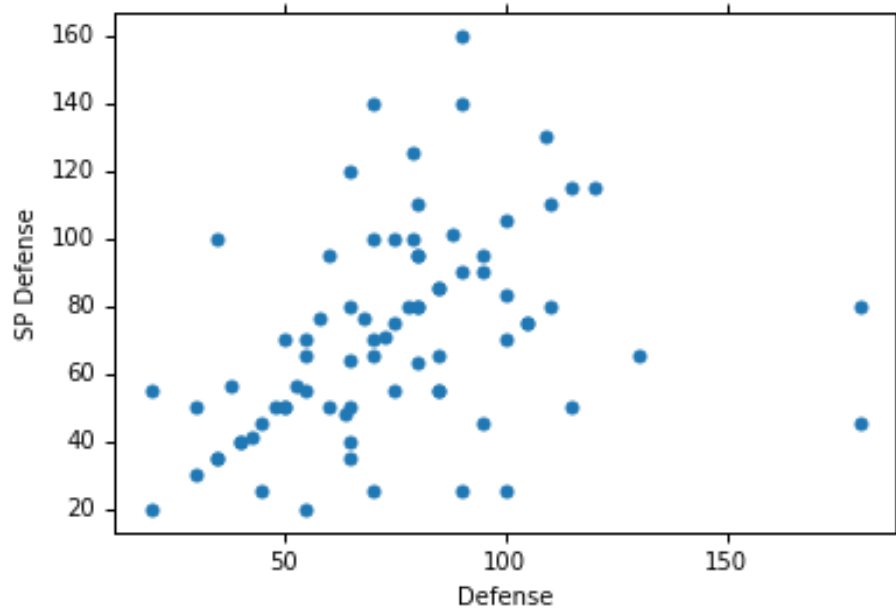$$\mu^* = \frac{1}{79}\sum_{n=1}^{79} x^n \qquad \Sigma^* = \frac{1}{79}\sum_{n=1}^{79}(x^n - \mu^*)(x^n - \mu^*)^T$$
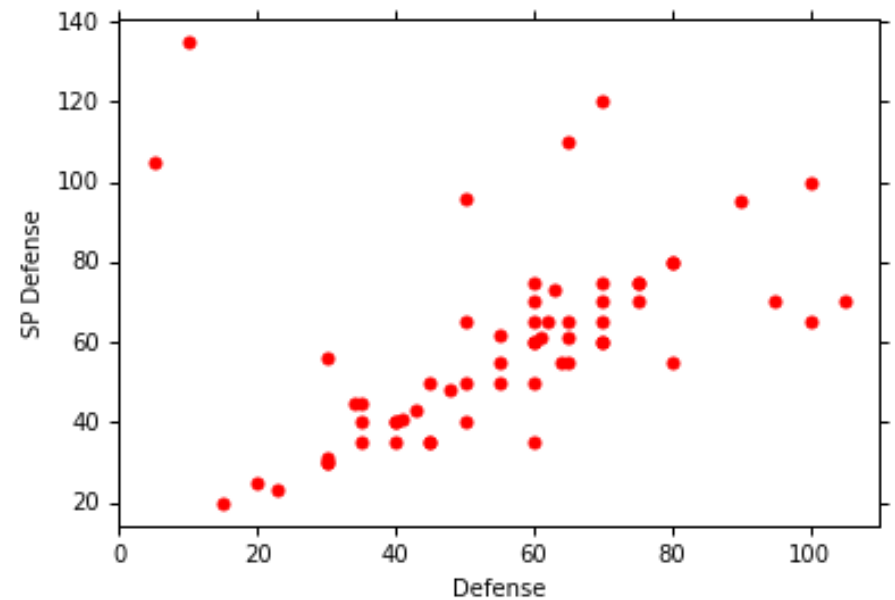
average

f(平均值)為最大值

Maximum Likelihood

Class 1: Water

Class 2: Normal

$$\mu^1 = \begin{bmatrix} 75.0 \\ 71.3 \end{bmatrix} \quad \Sigma^1 = \begin{bmatrix} 874 & 327 \\ 327 & 929 \end{bmatrix}$$

$$\mu^2 = \begin{bmatrix} 55.6 \\ 59.8 \end{bmatrix} \quad \Sigma^2 = \begin{bmatrix} 847 & 422 \\ 422 & 685 \end{bmatrix}$$

Now we can do classification ☺

$$f_{\mu^1,\Sigma^1}(x) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma^1|^{1/2}} exp\left\{-\frac{1}{2}(x-\mu^1)^T(\Sigma^1)^{-1}(x-\mu^1)\right\}$$

P(C1)
= 79 / (79 + 61) =0.56

$$\mu^1 = \begin{bmatrix} 75.0 \\ 71.3 \end{bmatrix} \quad \Sigma^1 = \begin{bmatrix} 874 & 327 \\ 327 & 929 \end{bmatrix}$$

根據79隻寶可夢的分數算出來的
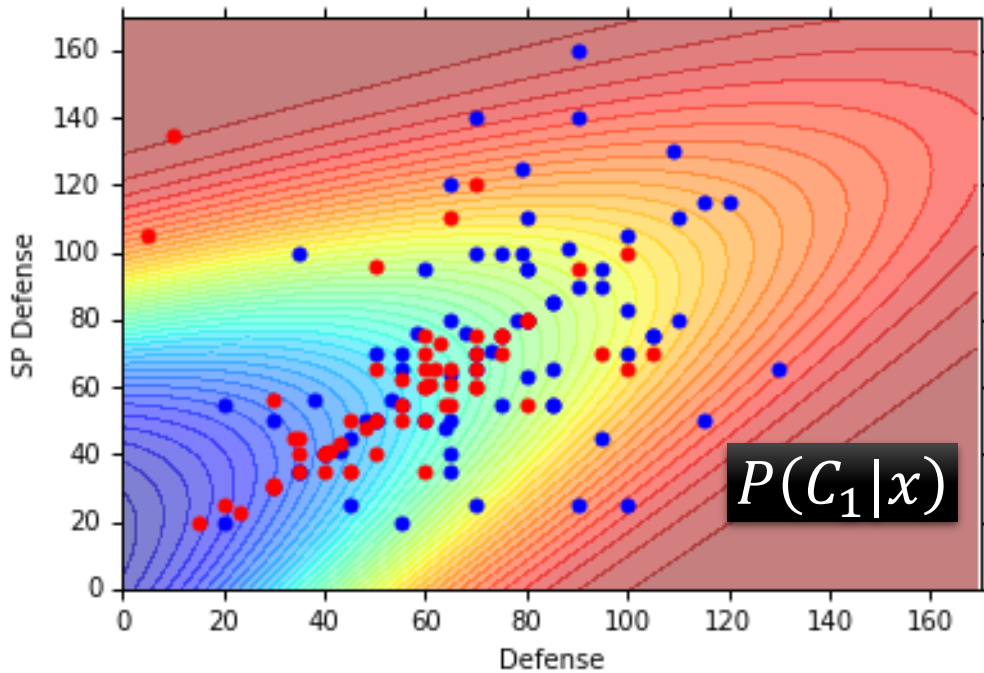
$$P(C_1|x) = \frac{P(x|C_1)P(C_1)}{P(x|C_1)P(C_1) + P(x|C_2)P(C_2)}$$

$$f_{\mu^2,\Sigma^2}(x) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma^2|^{1/2}} exp\left\{-\frac{1}{2}(x-\mu^2)^T(\Sigma^2)^{-1}(x-\mu^2)\right\}$$

P(C2)
= 61 / (79 + 61)
=0.44

$$\mu^2 = \begin{bmatrix} 55.6 \\ 59.8 \end{bmatrix} \quad \Sigma^2 = \begin{bmatrix} 847 & 422 \\ 422 & 685 \end{bmatrix}$$

If $P(C_1|x) > 0.5$ ➡ x belongs to class 1 (Water)

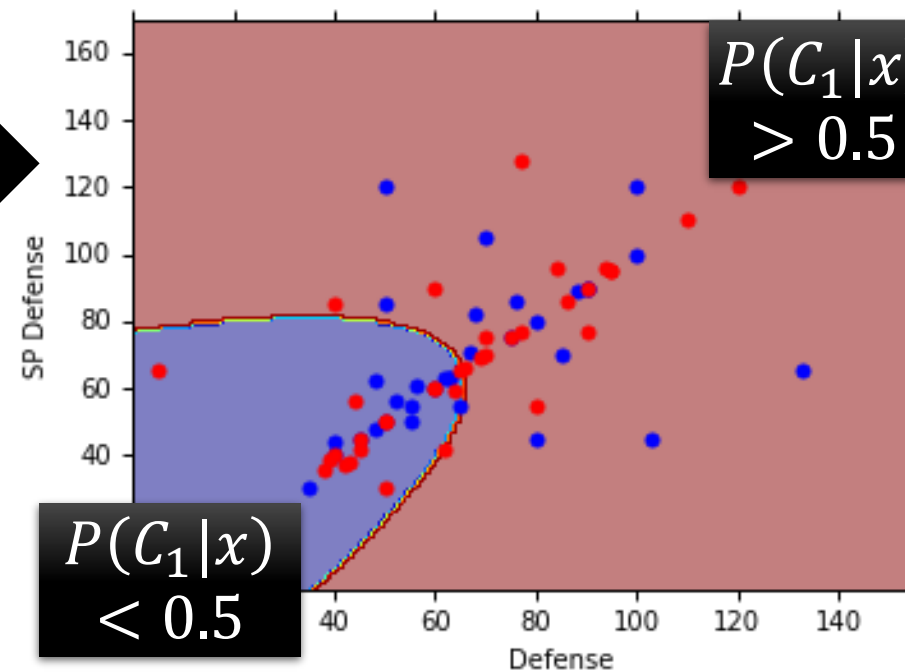Blue points: C_1 (Water), Red points: C_2 (Normal)

How's the results?

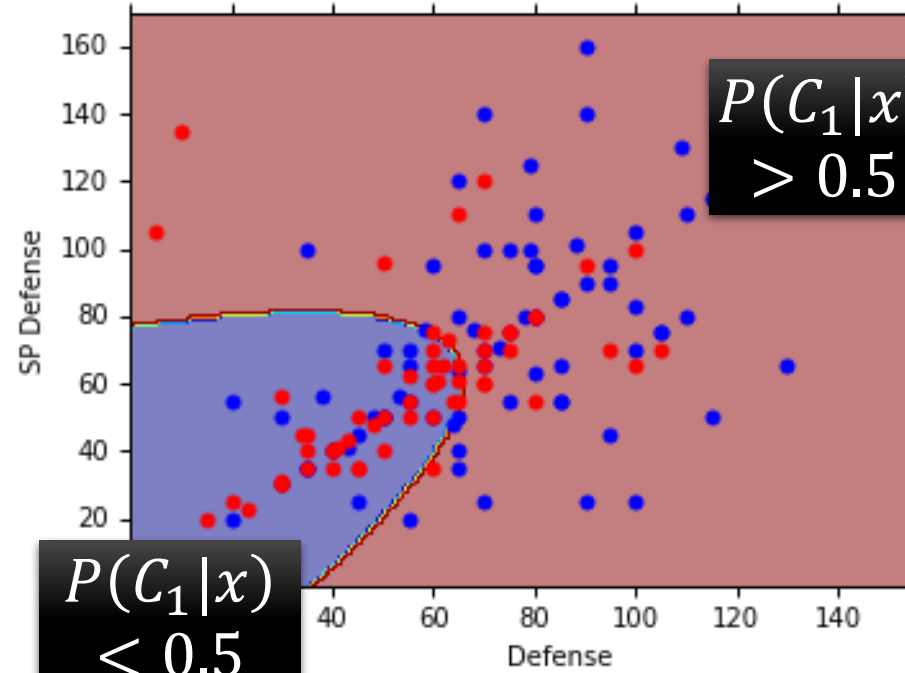Testing data: 47% accuracy ➡

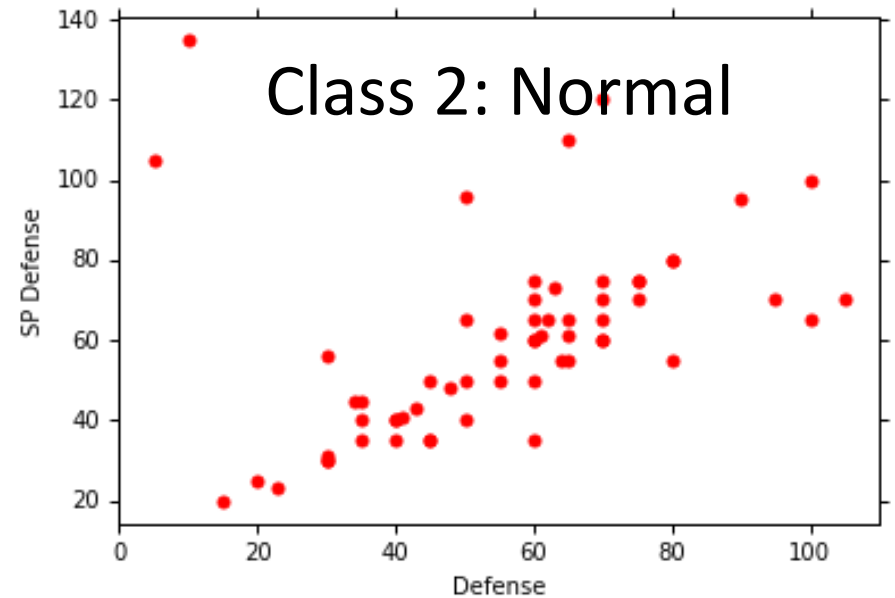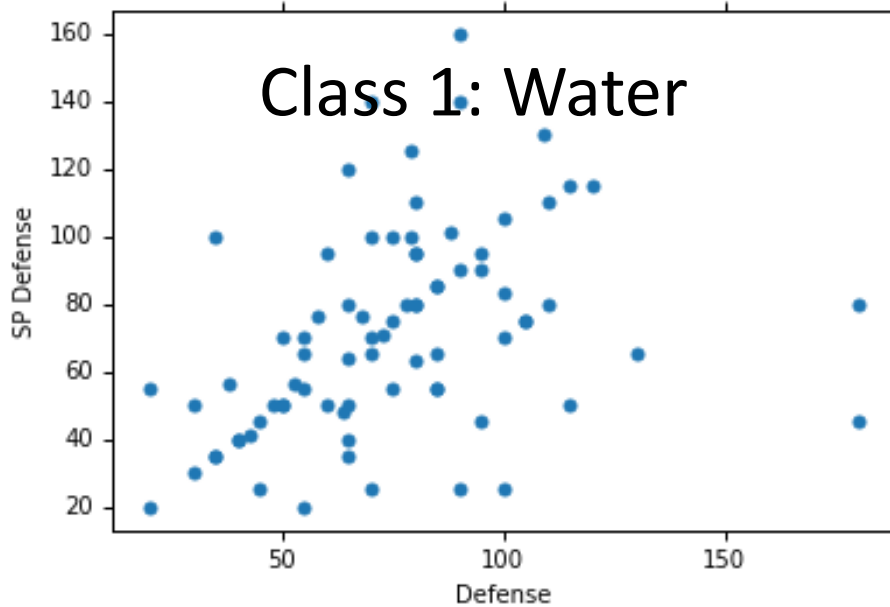All: total, hp, att, sp att,

de, sp de, speed (7 features)

μ^1, μ^2: 7-dim vector

Σ^1, Σ^2: 7 x 7 matrices

54% accuracy ... ☹

$P(C_1|x)$

$P(C_1|x) > 0.5$

$P(C_1|x) < 0.5$

$P(C_1|x) > 0.5$

$P(C_1|x) < 0.5$

Modifying Model

Class 1: Water

Class 2: Normal

$$\mu^1 = \begin{bmatrix} 75.0 \\ 71.3 \end{bmatrix} \quad \Sigma^1 = \begin{bmatrix} 874 & 327 \\ 327 & 929 \end{bmatrix} \qquad \mu^2 = \begin{bmatrix} 55.6 \\ 59.8 \end{bmatrix} \quad \Sigma^2 = \begin{bmatrix} 847 & 422 \\ 422 & 685 \end{bmatrix}$$

The same Σ

使用共同的gaussian

Less parameters

共用variance，減少parameters避免overfitting

Modifying Model

Ref: Bishop, chapter 4.2.2

- Maximum likelihood

"Water" type Pokémons:

$$x^1, x^2, x^3, \ldots\ldots, x^{79}$$

"Normal" type Pokémons:

$$x^{80}, x^{81}, x^{82}, \ldots\ldots, x^{140}$$

μ^1 79筆平均 Σ 61筆平均 μ^2
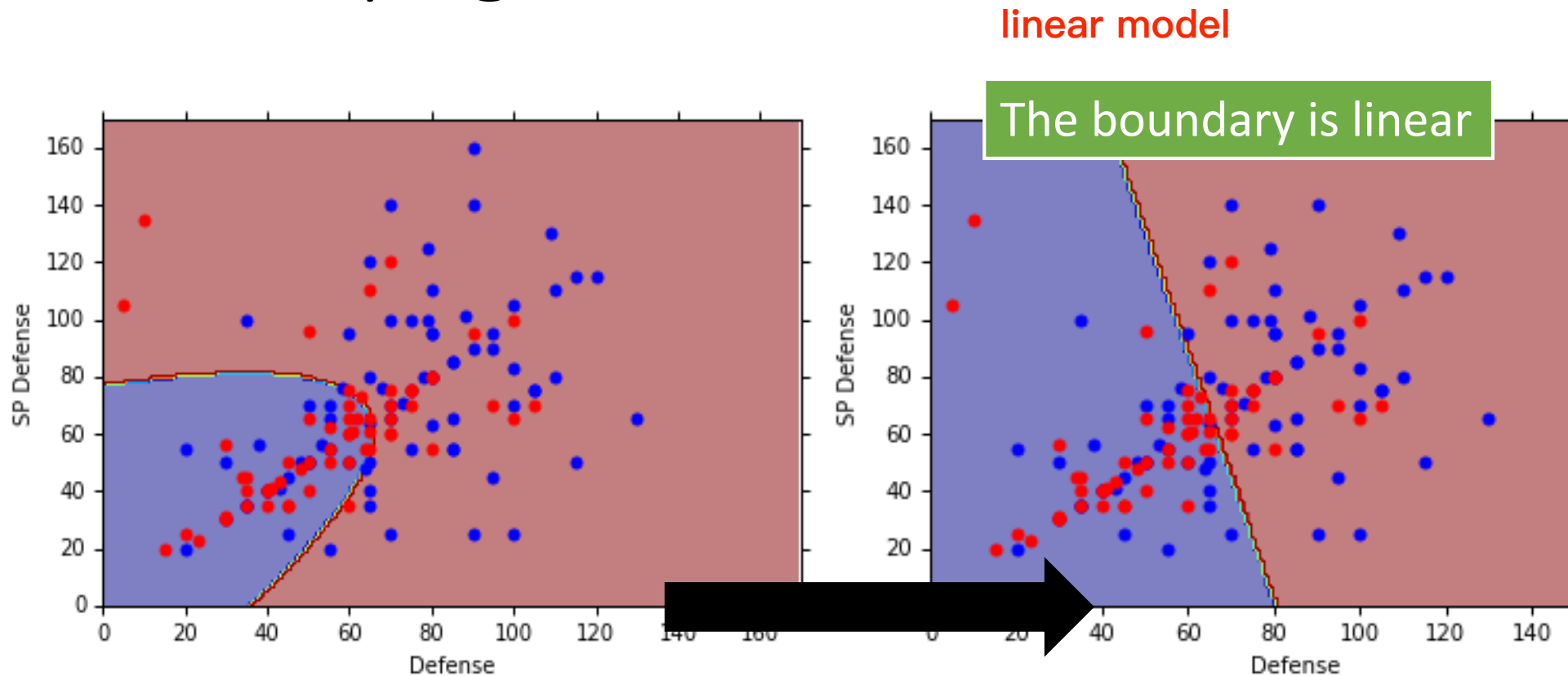
Find μ^1, μ^2, Σ maximizing the likelihood $L(\mu^1, \mu^2, \Sigma)$

$$L(\mu^1, \mu^2, \Sigma) = f_{\mu^1, \Sigma}(x^1) f_{\mu^1, \Sigma}(x^2) \cdots f_{\mu^1, \Sigma}(x^{79})$$
$$\times f_{\mu^2, \Sigma}(x^{80}) f_{\mu^2, \Sigma}(x^{81}) \cdots f_{\mu^2, \Sigma}(x^{140})$$

μ^1 and μ^2 is the same $\Sigma = \dfrac{79}{140}\Sigma^1 + \dfrac{61}{140}\Sigma^2$

Modifying Model

linear model

The boundary is linear

The same covariance matrix

All: total, hp, att, sp att, de, sp de, speed

54% accuracy ➡ 73% accuracy

共用covairance matrix

Three Steps

- Function Set (Model):

$$P(C_1|x) = \frac{P(x|C_1)P(C_1)}{P(x|C_1)P(C_1) + P(x|C_2)P(C_2)}$$
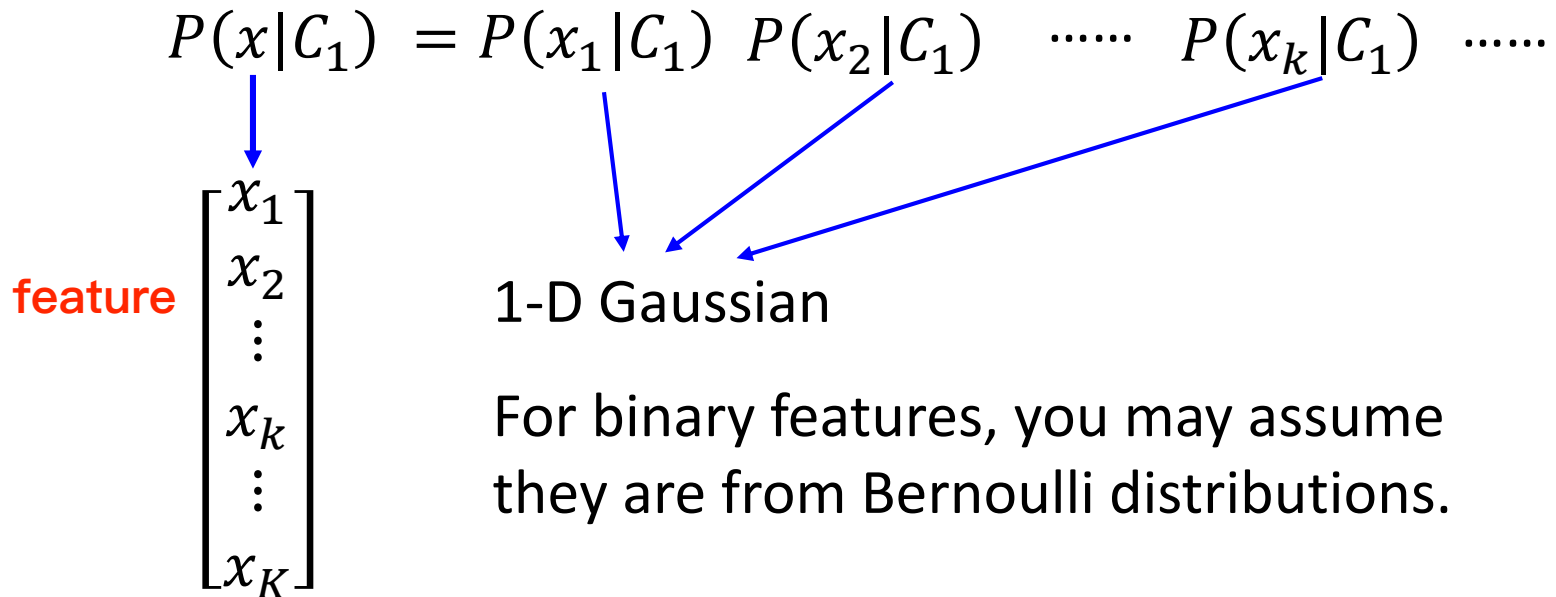
x ➡

If $P(C_1|x) > 0.5$, output: class 1

Otherwise, output: class 2

- Goodness of a function:
  - The mean μ and covariance Σ that maximizing the likelihood (the probability of generating data)
- Find the best function: easy

Probability Distribution

- You can always use the distribution you like ☺

$$P(x|C_1) = P(x_1|C_1) \; P(x_2|C_1) \quad \cdots\cdots \quad P(x_k|C_1) \quad \cdots\cdots$$

feature $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \\ \vdots \\ x_K \end{bmatrix}$

1-D Gaussian

For binary features, you may assume they are from Bernoulli distributions.

If you assume all the dimensions are independent, then you are using *Naive Bayes Classifier*.
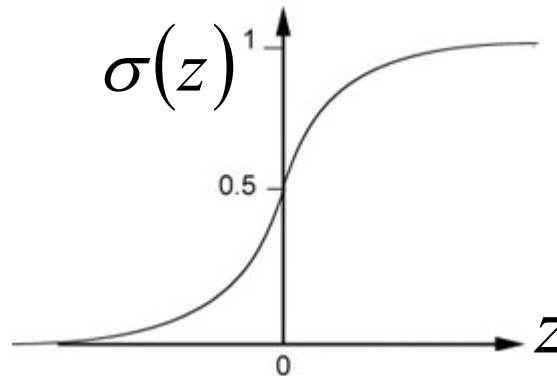
Posterior Probability

$$P(C_1|x) = \frac{P(x|C_1)P(C_1)}{P(x|C_1)P(C_1) + P(x|C_2)P(C_2)}$$

$$= \frac{1}{1 + \dfrac{P(x|C_2)P(C_2)}{P(x|C_1)P(C_1)}} \qquad = \frac{1}{1 + exp(-z)} \qquad = \sigma(z)$$

Sigmoid function

$$z = ln\frac{P(x|C_1)P(C_1)}{P(x|C_2)P(C_2)}$$

$\sigma(z)$

1

0.5

0

z

Warning of Math

Posterior Probability

$$P(C_1|x) = \sigma(z)$$ sigmoid $$z = ln\frac{P(x|C_1)P(C_1)}{P(x|C_2)P(C_2)}$$

class1在C1中出現的機率

$$z = ln\frac{P(x|C_1)}{P(x|C_2)} + ln\frac{P(C_1)}{P(C_2)} \implies \frac{\frac{N_1}{N_1+N_2}}{\frac{N_2}{N_1+N_2}} = \frac{N_1}{N_2}$$

class2在C2中出後的機率

$$P(x|C_1) = \frac{1}{(2\pi)^{D/2}}\frac{1}{|\Sigma^1|^{1/2}} exp\left\{-\frac{1}{2}(x-\mu^1)^T(\Sigma^1)^{-1}(x-\mu^1)\right\}$$

$$P(x|C_2) = \frac{1}{(2\pi)^{D/2}}\frac{1}{|\Sigma^2|^{1/2}} exp\left\{-\frac{1}{2}(x-\mu^2)^T(\Sigma^2)^{-1}(x-\mu^2)\right\}$$

$$z = \boxed{ln \frac{P(x|C_1)}{P(x|C_2)}} + ln \boxed{\frac{P(C_1)}{P(C_2)}} = \frac{N_1}{N_2}$$

$$P(x|C_1) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma^1|^{1/2}} exp\left\{-\frac{1}{2}(x-\mu^1)^T(\Sigma^1)^{-1}(x-\mu^1)\right\}$$

$$P(x|C_2) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma^2|^{1/2}} exp\left\{-\frac{1}{2}(x-\mu^2)^T(\Sigma^2)^{-1}(x-\mu^2)\right\}$$

$$ln \frac{\cancel{\frac{1}{(2\pi)^{D/2}}} \boxed{\frac{1}{|\Sigma^1|^{1/2}}} \boxed{exp\left\{-\frac{1}{2}(x-\mu^1)^T(\Sigma^1)^{-1}(x-\mu^1)\right\}}}{\cancel{\frac{1}{(2\pi)^{D/2}}} \boxed{\frac{1}{|\Sigma^2|^{1/2}}} \boxed{exp\left\{-\frac{1}{2}(x-\mu^2)^T(\Sigma^2)^{-1}(x-\mu^2)\right\}}}$$

$$= \boxed{ln \frac{|\Sigma^2|^{1/2}}{|\Sigma^1|^{1/2}}} \boxed{exp\left\{-\frac{1}{2}[(x-\mu^1)^T(\Sigma^1)^{-1}(x-\mu^1) \\ - (x-\mu^2)^T(\Sigma^2)^{-1}(x-\mu^2)]\right\}}$$

$$= ln \frac{|\Sigma^2|^{1/2}}{|\Sigma^1|^{1/2}} - \frac{1}{2}[(x-\mu^1)^T(\Sigma^1)^{-1}(x-\mu^1) - (x-\mu^2)^T(\Sigma^2)^{-1}(x-\mu^2)]$$

$$z = \boxed{ln\frac{P(x|C_1)}{P(x|C_2)}} + ln\boxed{\frac{P(C_1)}{P(C_2)}} = \frac{N_1}{N_2}$$

$$= ln\frac{|\Sigma^2|^{1/2}}{|\Sigma^1|^{1/2}} - \frac{1}{2}\left[\underline{(x-\mu^1)^T(\Sigma^1)^{-1}(x-\mu^1)} - \underline{(x-\mu^2)^T(\Sigma^2)^{-1}(x-\mu^2)}\right]$$

$$(x-\mu^1)^T(\Sigma^1)^{-1}(x-\mu^1)$$

$$= x^T(\Sigma^1)^{-1}x \underline{- x^T(\Sigma^1)^{-1}\mu^1 - (\mu^1)^T(\Sigma^1)^{-1}x} + (\mu^1)^T(\Sigma^1)^{-1}\mu^1$$

$$= x^T(\Sigma^1)^{-1}x \underline{- 2(\mu^1)^T(\Sigma^1)^{-1}x} + (\mu^1)^T(\Sigma^1)^{-1}\mu^1$$

$$(x-\mu^2)^T(\Sigma^2)^{-1}(x-\mu^2)$$

$$= x^T(\Sigma^2)^{-1}x - 2(\mu^2)^T(\Sigma^2)^{-1}x + (\mu^2)^T(\Sigma^2)^{-1}\mu^2$$

$$z = ln\frac{|\Sigma^2|^{1/2}}{|\Sigma^1|^{1/2}} - \frac{1}{2}x^T(\Sigma^1)^{-1}x + (\mu^1)^T(\Sigma^1)^{-1}x - \frac{1}{2}(\mu^1)^T(\Sigma^1)^{-1}\mu^1$$

推倒結果

$$+ \frac{1}{2}x^T(\Sigma^2)^{-1}x - (\mu^2)^T(\Sigma^2)^{-1}x + \frac{1}{2}(\mu^2)^T(\Sigma^2)^{-1}\mu^2 + ln\frac{N_1}{N_2}$$

End of Warning

$$P(C_1|x) = \sigma(z)$$

$$z = ln\frac{|\Sigma^2|^{1/2}}{|\Sigma^1|^{1/2}} \; \cancel{-\frac{1}{2}x^T(\Sigma^1)^{-1}x} + (\mu^1)^T(\Sigma^1)^{-1}x - \frac{1}{2}(\mu^1)^T(\Sigma^1)^{-1}\mu^1$$

$$+ \cancel{\frac{1}{2}x^T(\Sigma^2)^{-1}x} - (\mu^2)^T(\Sigma^2)^{-1}x + \frac{1}{2}(\mu^2)^T(\Sigma^2)^{-1}\mu^2 + ln\frac{N_1}{N_2}$$

$$\Sigma_1 = \Sigma_2 = \Sigma$$

$$z = \underbrace{(\mu^1 - \mu^2)^T\Sigma^{-1}}_{\boldsymbol{w^T} \text{ vector}}x \underbrace{- \frac{1}{2}(\mu^1)^T\Sigma^{-1}\mu^1 + \frac{1}{2}(\mu^2)^T\Sigma^{-1}\mu^2 + ln\frac{N_1}{N_2}}_{\text{b scalor}}$$

$$P(C_1|x) = \sigma(w \cdot x + b)$$

How about directly find **w** and b?

In generative model, we estimate $N_1, N_2, \mu^1, \mu^2, \Sigma$

Gaussian

Then we have **w** and b

Reference

- Bishop: Chapter 4.1 – 4.2
- Data: https://www.kaggle.com/abcsds/pokemon
- Useful posts:
  - https://www.kaggle.com/nishantbhadauria/d/abcsds/pokemon/pokemon-speed-attack-hp-defense-analysis-by-type
  - https://www.kaggle.com/nikos90/d/abcsds/pokemon/mastering-pokebars/discussion
  - https://www.kaggle.com/ndrewgele/d/abcsds/pokemon/visualizing-pok-mon-stats-with-seaborn/discussion

Acknowledgment

- 感謝 江貫榮 同學發現課程網頁上的日期錯誤
- 感謝 范廷瀚 同學提供寶可夢的 domain knowledge
- 感謝 Victor Chen 發現投影片上的打字錯誤