Classification: Probabilistic Generative Model

Classification

- Credit Scoring
  - 输入特征 → Input: income, savings, profession, age, past financial history ……
  - 输出分类 ← Output: accept or refuse
- Medical Diagnosis
  - Input: current symptoms, age, gender, past medical history ……
  - Output: which kind of diseases
- Handwritten character recognition

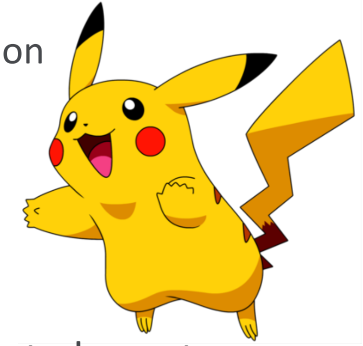
Input:  output: 金

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

Example Application

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

pokemon games (*NOT* pokemon cards or Pokemon Go)



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

攻擊方的屬性 / 防禦方的屬性

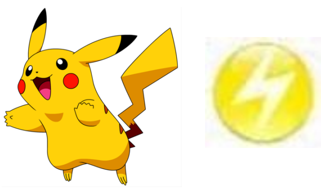
×	一般	格鬥	飛行	毒	地面	岩石	蟲	幽靈	鋼	火	水	草	電	超能力	冰	龍	惡	妖精
一般	1×	1×	1×	1×	1×	½×	1×	0×	½×	1×	1×	1×	1×	1×	1×	1×	1×	1×
格鬥	2×	1×	½×	½×	1×	2×	½×	0×	2×	1×	1×	1×	1×	½×	2×	1×	2×	½×
飛行	1×	2×	1×	1×	1×	½×	2×	1×	½×	1×	1×	2×	½×	1×	1×	1×	1×	1×
毒	1×	1×	1×	½×	½×	½×	1×	½×	0×	1×	1×	2×	1×	1×	1×	1×	1×	2×
地面	1×	1×	0×	2×	1×	2×	½×	1×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

這些倍數適用於XY及之後的遊戲。

How to do Classification

- Training data for Classification

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

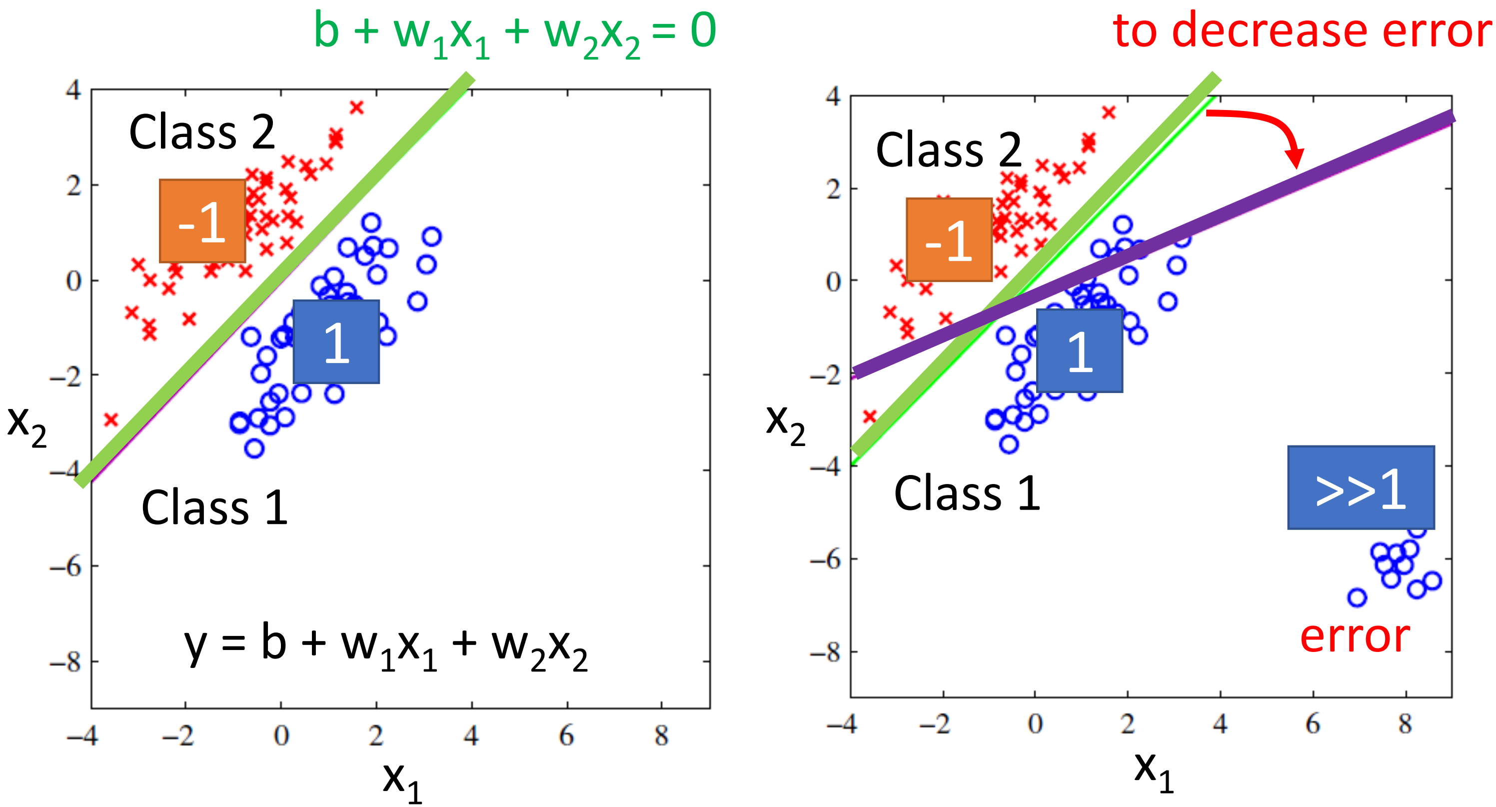


Classification as Regression?

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

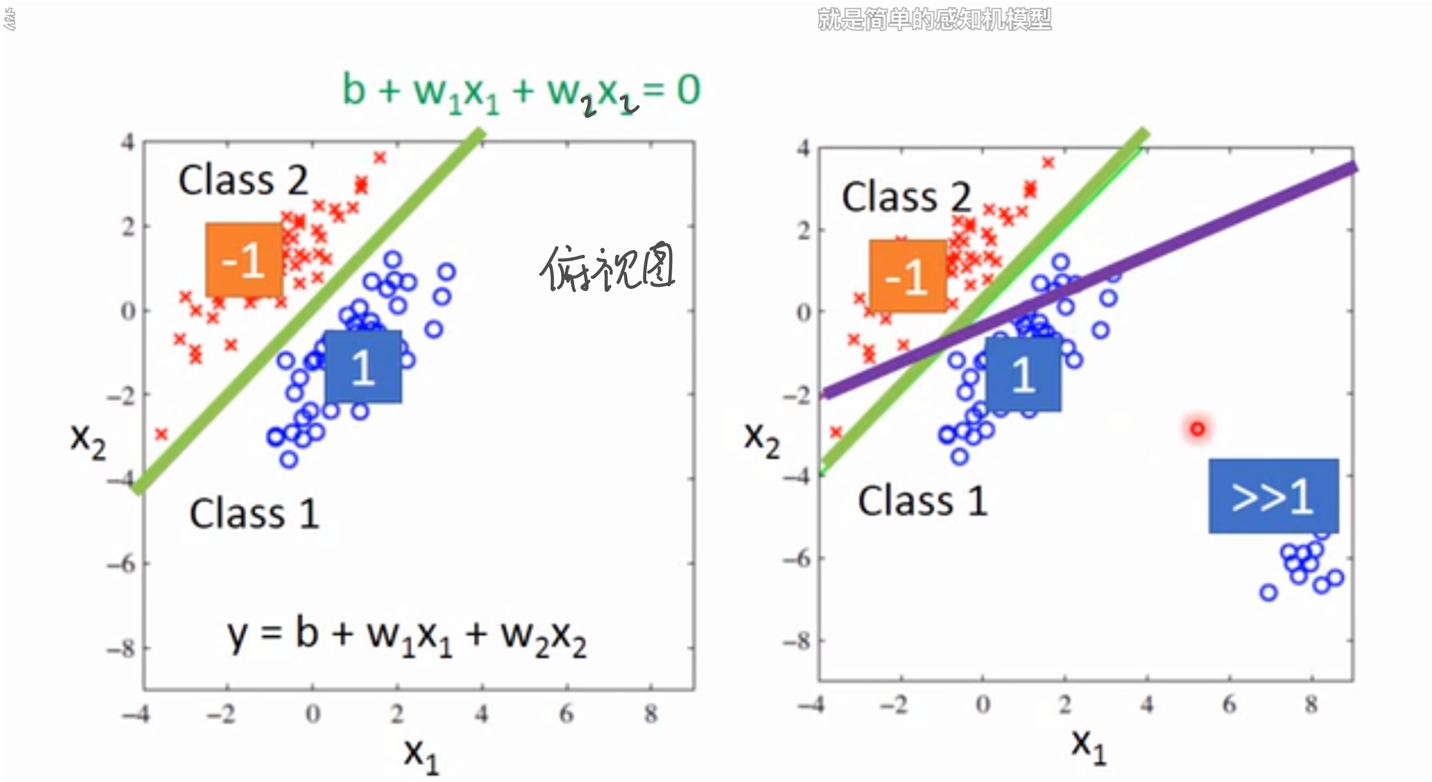


Penalize to the examples that are "too correct" … (Bishop, P186)

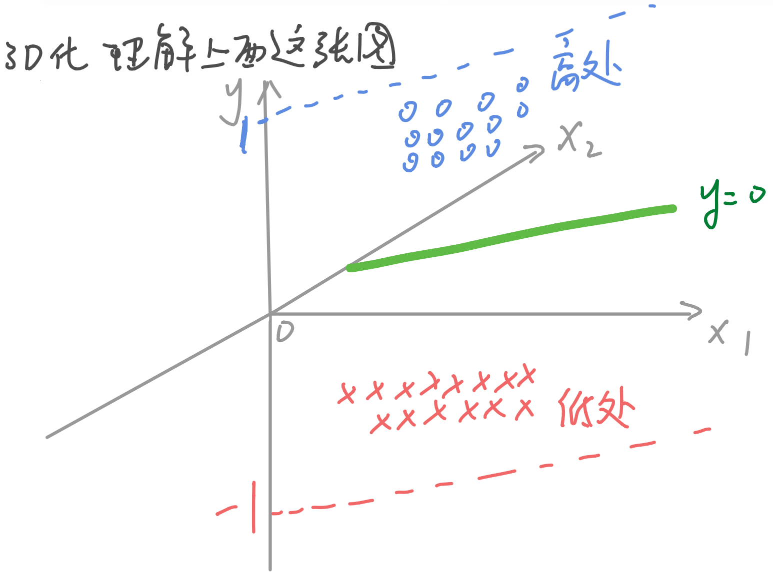
- Multiple class: Class 1 means the target is 1; Class 2 means the target is 2; Class 3 means the target is 3 …… problematic

一些问题通过线性回归无法解决(需要预测的不是某个数值)
(而是人为分出的类型)

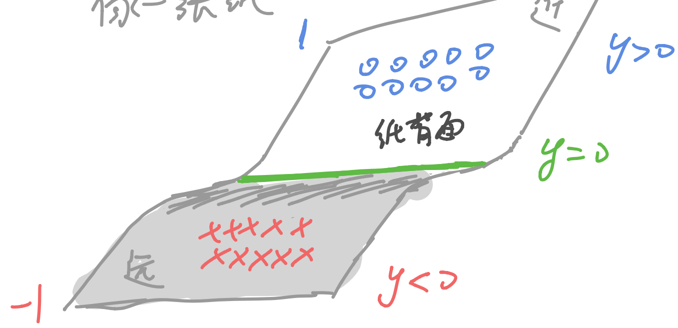
寻找一个方法 $\hat{f}$, input 为 $\vec{x}$, output 为 class1, class2 $\Rightarrow \hat{f}(\vec{x}) \Rightarrow \begin{cases} \text{class1} \\ \text{class2} \end{cases}$



3D化 理解上面这张图



像一张纸



Ideal Alternatives

- Function (Model):

$f(x)$

$x \rightarrow$

$g(x) > 0 \Rightarrow$ Output = class 1

$else$ $g(x) < 0 \Rightarrow$ Output = class 2

- Loss function:

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

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

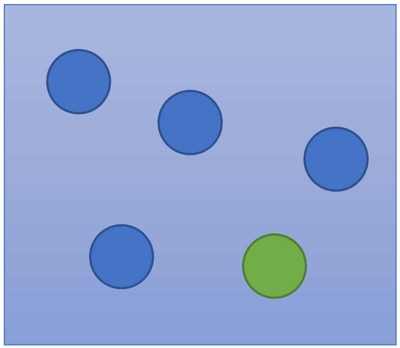
- Find the best function:
  - Example: Perceptron, SVM

Not Today

Two Boxes

Box 1

$P(B_1) = 2/3$

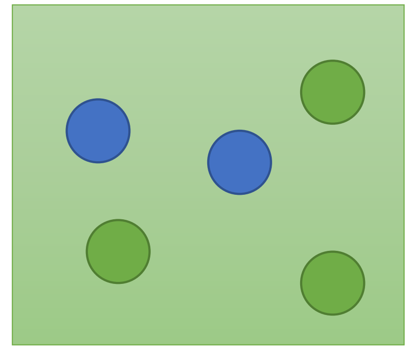


$P(\text{Blue}|B_1) = 4/5$

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

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?

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

先考虑这样一件事，从2个盒子中任选1个盒子，取一次球，若取出的是蓝球，问从Box1中取出的概率

Two Boxes

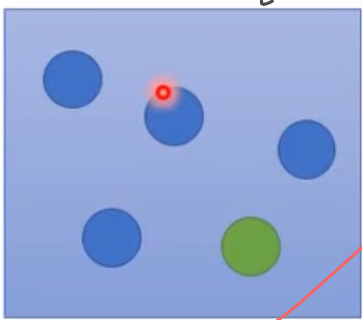
$B_i = \{$从第i个盒子取球$\}$

$A_i = \{$从第i个盒子取出的是蓝球$\}$

Box 1

$P(B_1) = \frac{1}{2}$

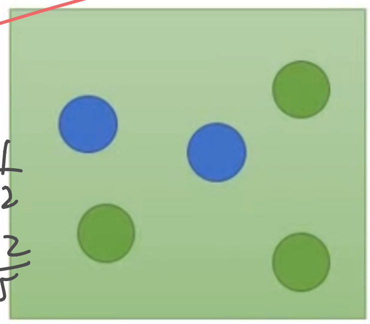
$P(A_1) = P(A|B_1) = \frac{4}{5}$



Box 2

$P(B_2) = \frac{1}{2}$

$P(A_2) = P(A|B_2) = \frac{2}{5}$



$$= P(B_1|A) = \frac{P(B_1)P(A|B_1)}{P(A)} = \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2)}$$

$$= \frac{P(B_1)\cdot P(A_1)}{P(B_1)\cdot P(A_1) + P(B_2)\cdot P(A_2)}$$

$$= \frac{\frac{1}{2}\cdot\frac{4}{5}}{\frac{1}{2}\cdot\frac{4}{5} + \frac{1}{2}\cdot\frac{2}{5}}$$

$$= \frac{\frac{4}{10}}{\frac{6}{10}} = \frac{2}{3}$$

 from one of the boxes

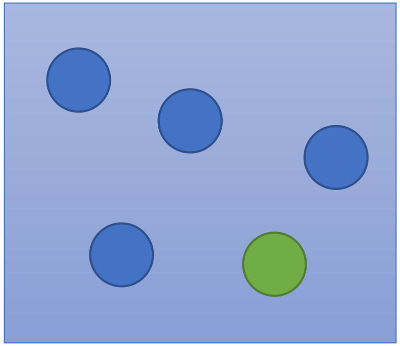
Where does it come from?

Two Classes

估计一 Estimating the Probabilities
From training data

Class 1

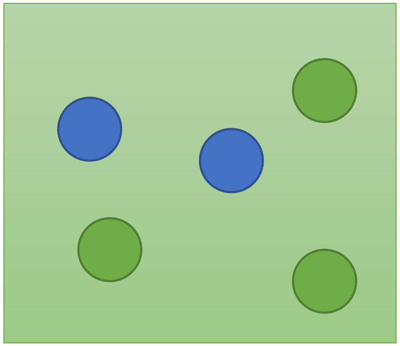
$P(C_1)$



$P(x|C_1)$

Class 2

$P(C_2)$



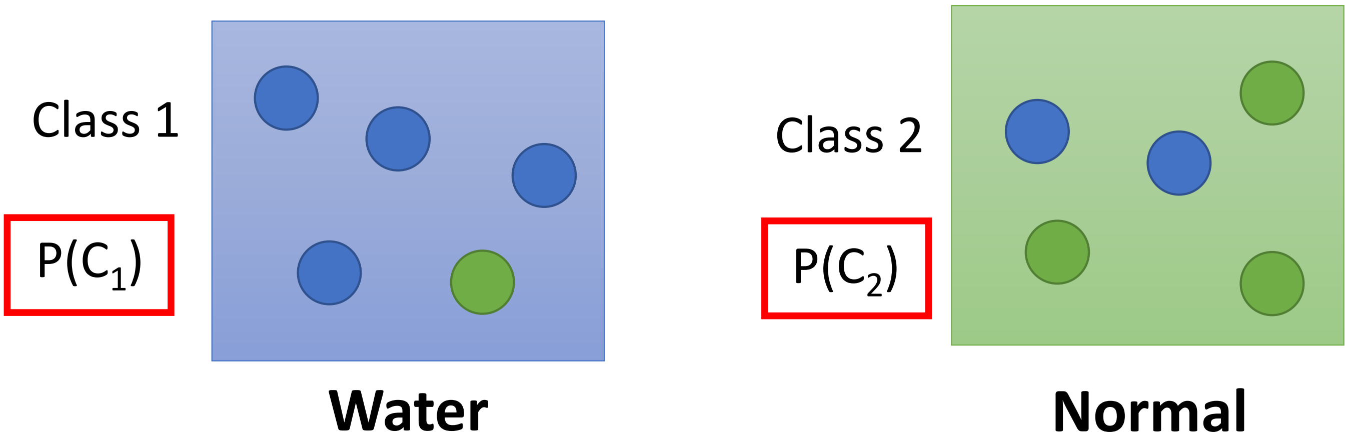
$P(x|C_2)$

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 $P(x) = P(x|C_1)P(C_1) + P(x|C_2)P(C_2)$

Prior



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) = ?

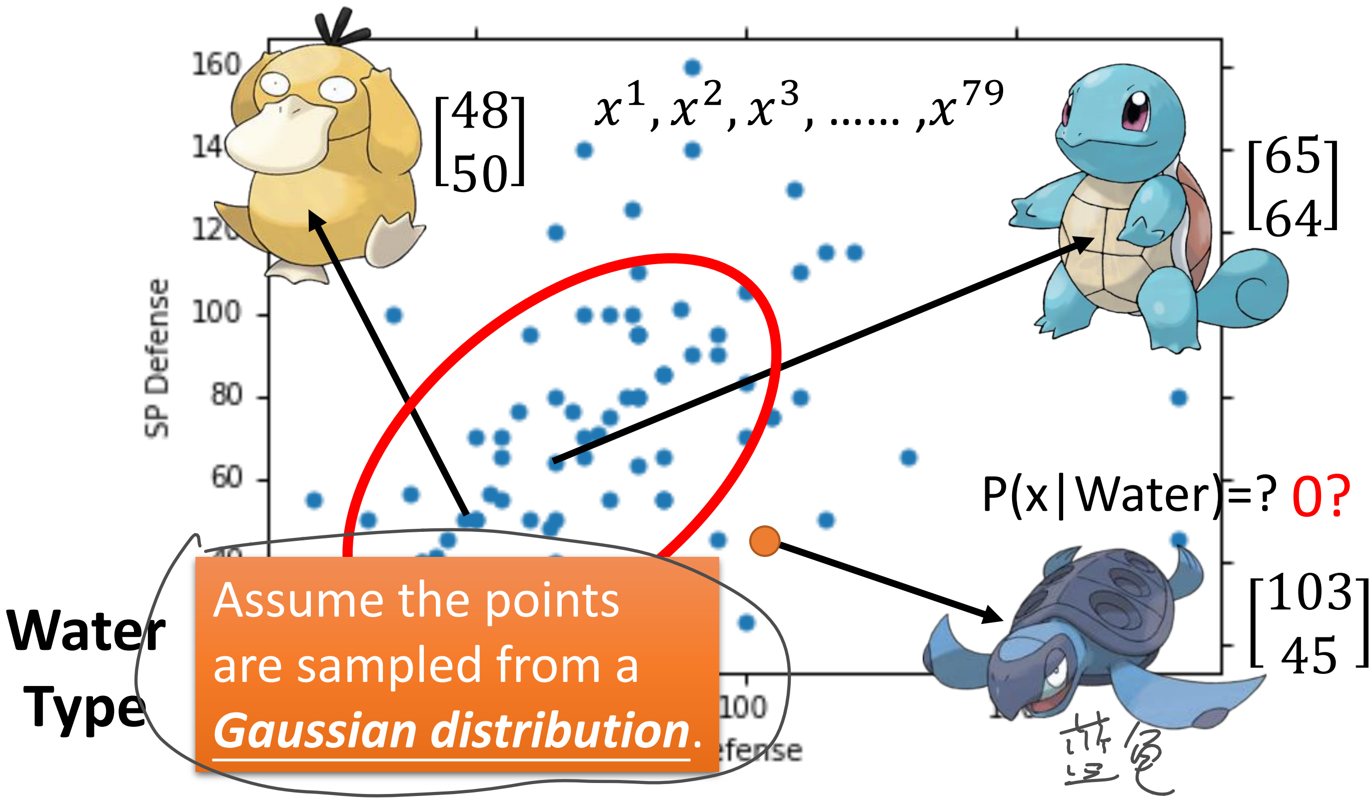
Each Pokémon is represented as a vector by its attribute. → feature

79 in total

Water Type

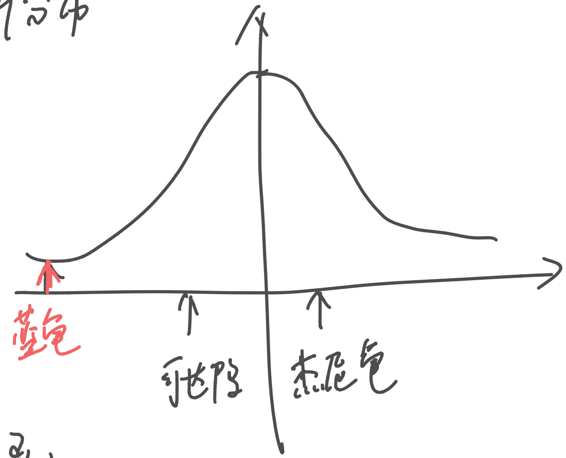
Probability from Class - Feature

- Considering **Defense** and **SP Defense**



每个样本被抽到的概率～高斯分布

高斯分布



被抽样的概率小

若求蓝龟的类 P(类?|蓝龟)，必须先知道 P(蓝龟|类?)

即 P(X蓝龟|Water)，蓝龟被抽到的概率问题，就变成了

确定这个高斯分布

Gaussian Distribution

exp{ c }表示 e^c，(e的c次幂)

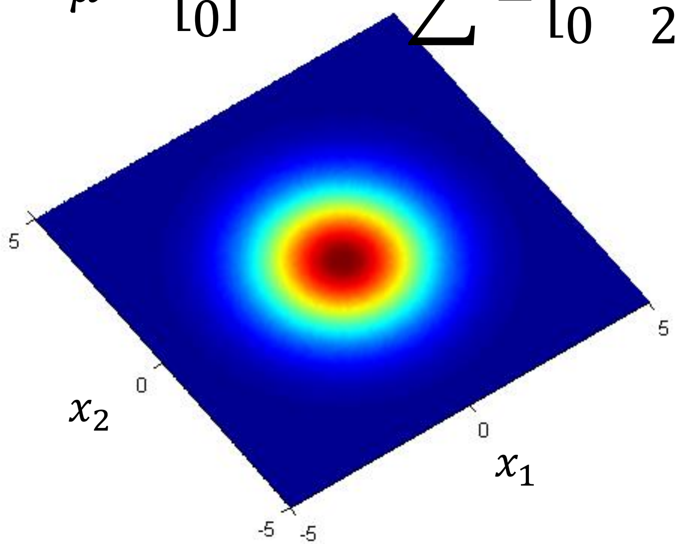
$$f_{\mu,\Sigma}(x) = \frac{1}{(2\pi)^{D/2}}\frac{1}{|\Sigma|^{1/2}}\boxed{exp}\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$$

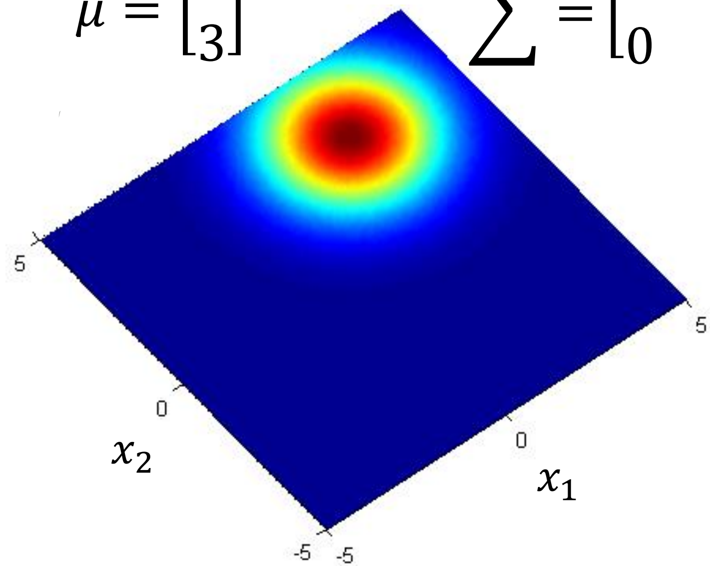
Input: vector x, output: probability of sampling x $\Sigma^{-1} = \frac{\Sigma^*}{|\Sigma|} = \frac{}{\sigma^2}$

The shape of the function determines by **mean $\boldsymbol{\mu}$** and **covariance matrix $\boldsymbol{\Sigma}$** μ，Σ 决定了高斯分布的形态（位置，坡度）

协方差

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





协方差矩阵 Covariance matrix

eg. $\Sigma = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$

① 样本 $x_1 \cdots x_n$

② 每个样本有2个维度 (a, b)

③ a与b协方差 $= \sigma(a,b) = \frac{1}{n-1}\sum_{i=1}^{n}(a_i-\bar{a})(b_i-\bar{b})$

④ 协方差矩阵 $\Sigma = \begin{pmatrix} \sigma(a,a) & \sigma(a,b) \\ \sigma(b,a) & \sigma(b,b) \end{pmatrix}$

eg. $\Sigma = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$

1. 方差和协方差的定义

> 在统计学中，**方差**是用来度量**单个随机变量**的**离散程度**，而协方差则一般用来刻画**两个随机变量**的**相似程度**，其中，**方差**的计算公式为
>
> $$\sigma_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$
>
> 其中，n 表示样本量，符号 $\bar{x}$ 表示观测样本的均值，这个定义在初中阶段就已经开始接触了。

在此基础上，**协方差**的计算公式被定义为

$$\sigma(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

在公式中，符号 $\bar{x},\bar{y}$ 分别表示两个随机变量所对应的观测样本均值，据此，我们发现：方差 σ_x^2 可视作随机变量 x 关于其自身的协方差 $\sigma(x,x)$.

2. 从方差/协方差到协方差矩阵

根据方差的定义，给定 d 个随机变量 $x_k, k=1,2,\ldots,d$ ，则这些**随机变量的方差**为

$$\sigma(x_k,x_k) = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ki}-\bar{x}_k)^2, k=1,2,\ldots,d$$

其中，为方便书写，x_{ki} 表示随机变量 x_k 中的第 i 个观测样本，n 表示样本量，每个随机变量所对应的观测样本数量均为 n 。

对于这些随机变量，我们还可以根据协方差的定义，求出**两两之间的协方差**，即

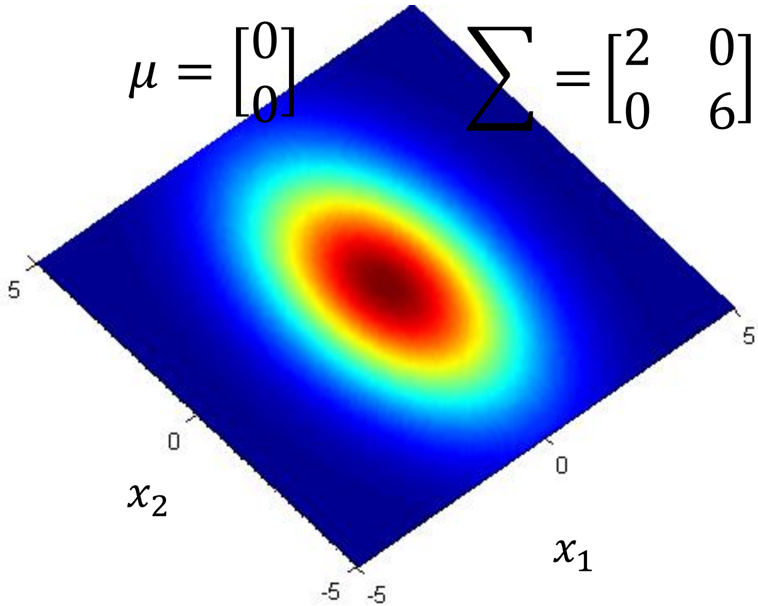
Gaussian Distribution

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

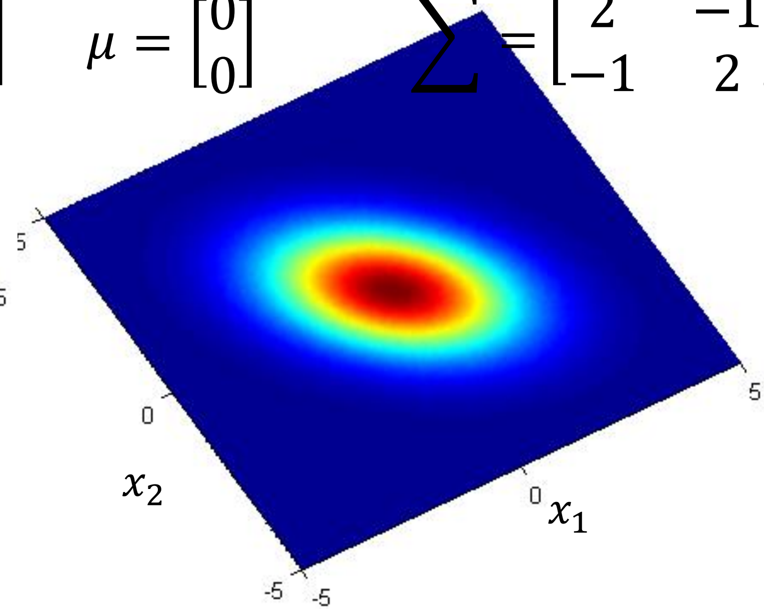
Input: vector x, output: probability of sampling x

The shape of the function determines by **mean** $\boldsymbol{\mu}$ and **covariance matrix** $\boldsymbol{\Sigma}$

$$\mu = \begin{bmatrix}0\\0\end{bmatrix} \qquad \sum = \begin{bmatrix}2 & 0\\0 & 6\end{bmatrix}$$



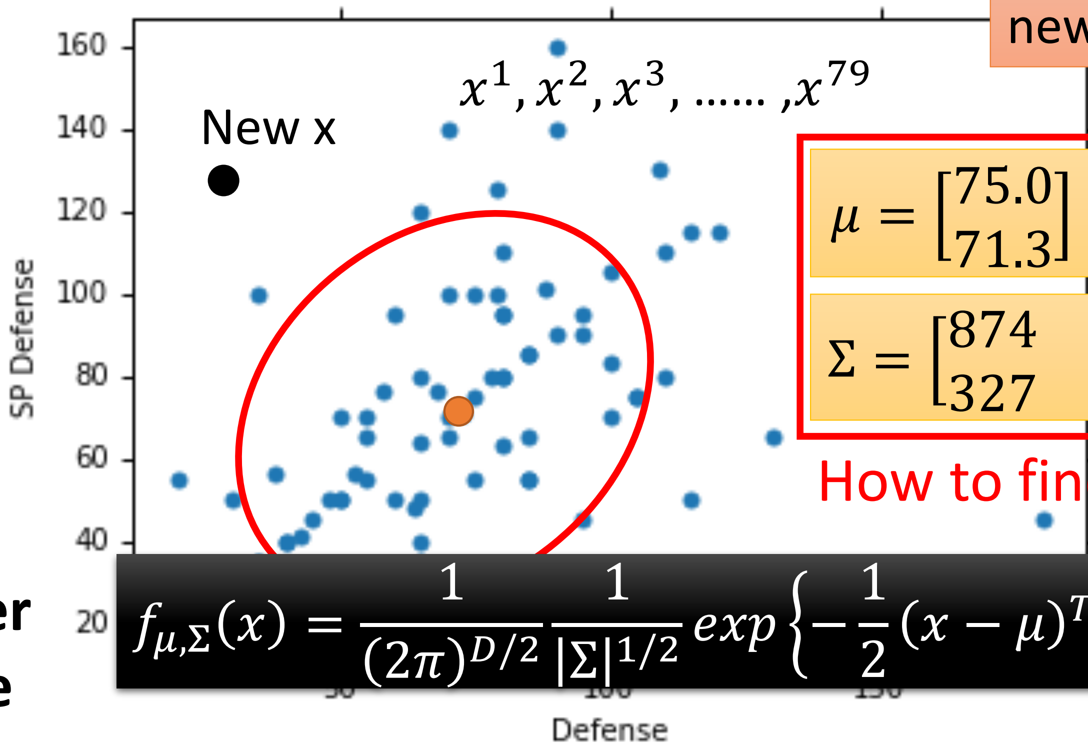
$$\mu = \begin{bmatrix}0\\0\end{bmatrix} \qquad \sum = \begin{bmatrix}2 & -1\\-1 & 2\end{bmatrix}$$



Probability from Class

Assume the points are sampled from a Gaussian distribution

Find the Gaussian distribution behind them → Probability for new points



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

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

How to find them?

Water Type

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

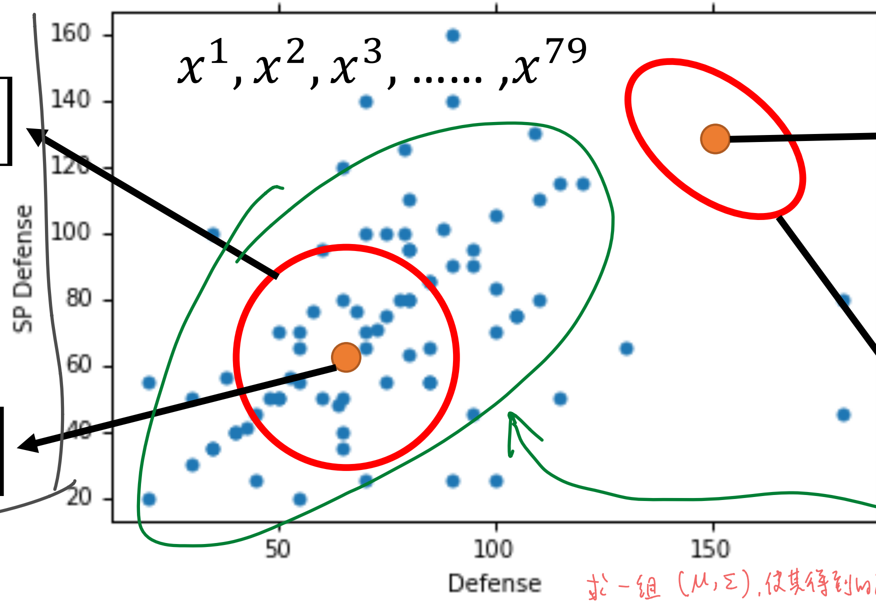
Maximum Likelihood

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

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

分布1

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



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

分布2

$$\Sigma = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

求一组 (μ,Σ)，使其得到的高斯分布，生成 $\boxed{\text{这堆 样本点}}$ 可能性最大

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

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

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***

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

Likelihood
不是Loss

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

$$\begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \\ \vdots \end{pmatrix}^T \rightarrow (x_1 - \mu_1, x_2 - \mu_2, \cdots)$$

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

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

average

不同的 μ 与 Σ 表了不同的高斯分布

我们要找到可能性最大的分布

高斯分布的密度函数有了 $f_{\mu,\Sigma}(x)$ (看上一页)

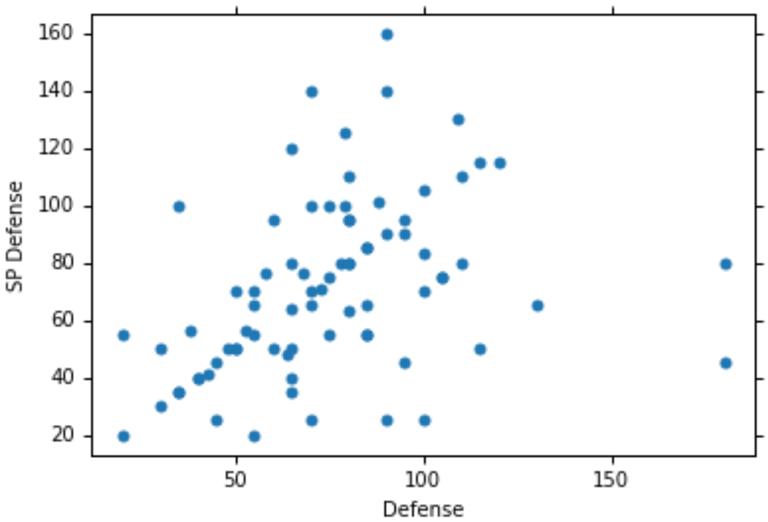
最大可能性函数 $L(\mu,\Sigma) = f_{\mu,\Sigma}(x^1) \cdot f_{\mu,\Sigma}(x^2) \cdots f_{\mu,\Sigma}(x^{79})$

(似然) (Likelihood)

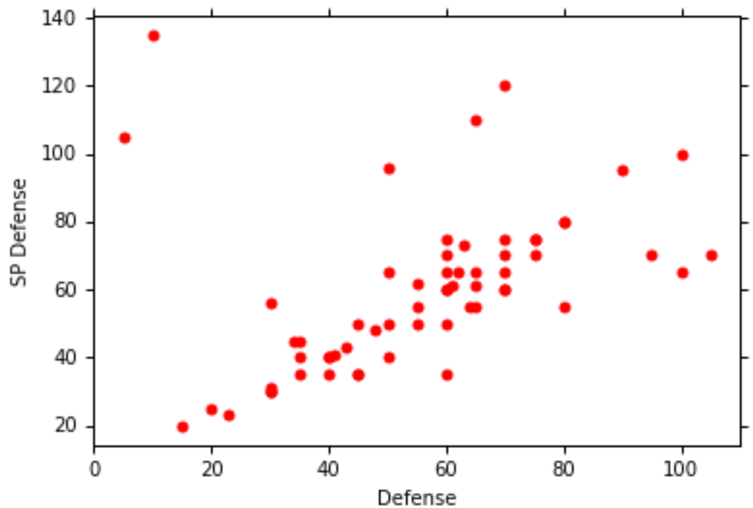
代入所有 $x^1 \cdots x^{79}$ 样本，找到一组 μ,Σ，使 L 最大

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 ☺

训练的结果 μ^1,Σ^1

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

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

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

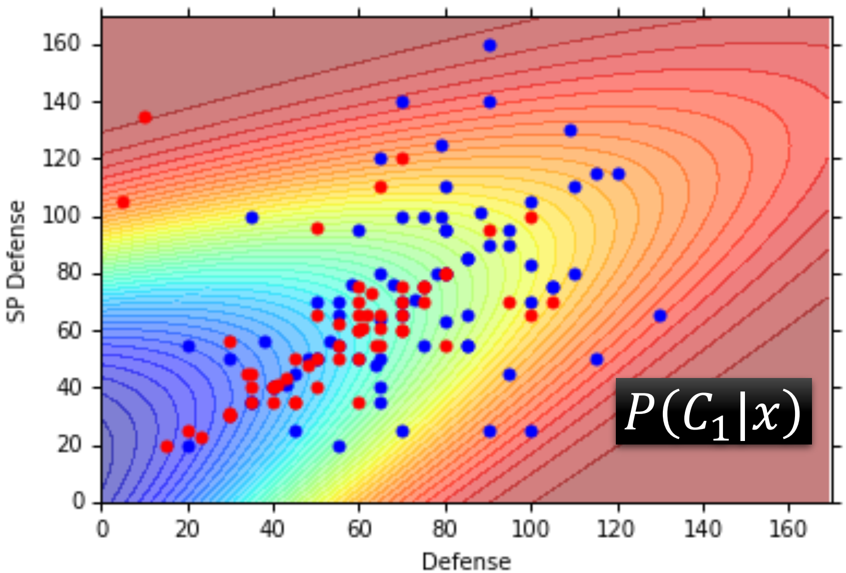
$$P(C_1|x) = \frac{P(x|C_1)P(C_1)}{P(x|C_1)P(C_1) + P(x|C_2)P(C_2)}$$

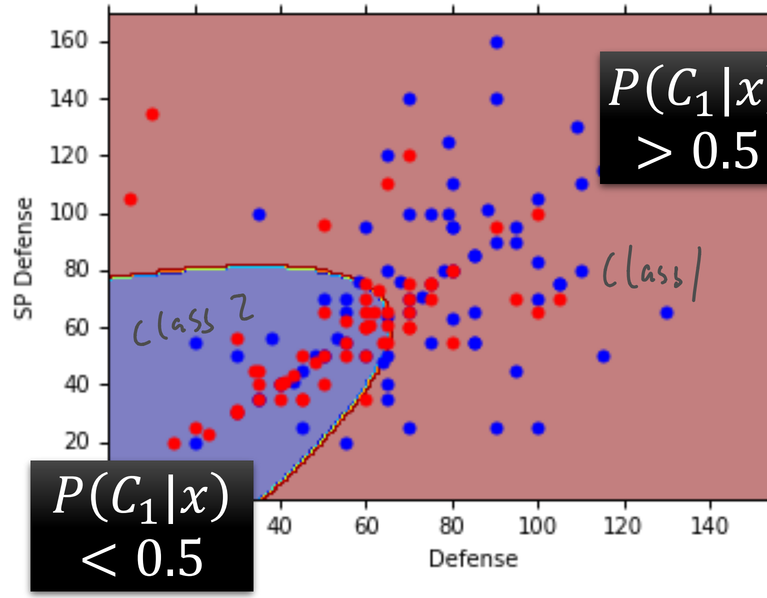
$$f_{\mu^2,\Sigma^2}(x) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma^2|^{1/2}} exp\left\{-\frac{1}{2}(x-\mu^2)^T(\Sigma^2)^{-1}(x-\mu^2)\right\}$$

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?

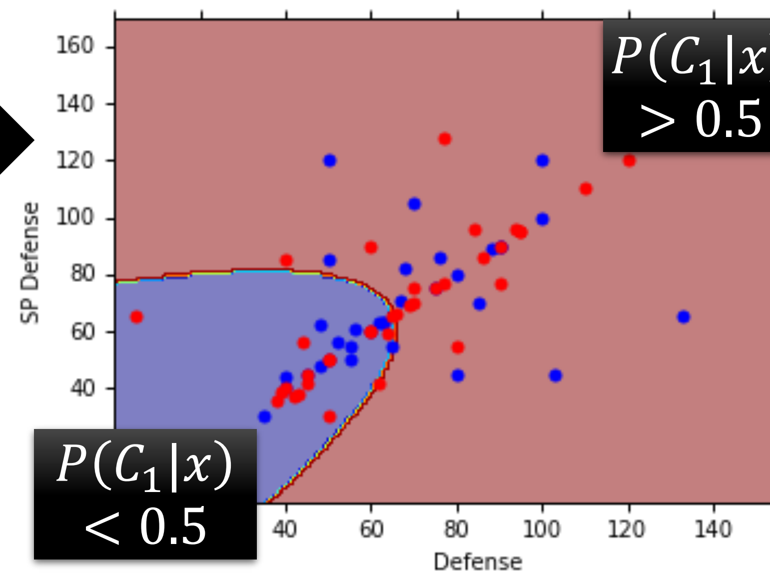
Testing data: 47% accuracy

All: total, hp, att, sp att, de, sp de, speed (7 features)

μ^1, μ^2: 7-dim vector

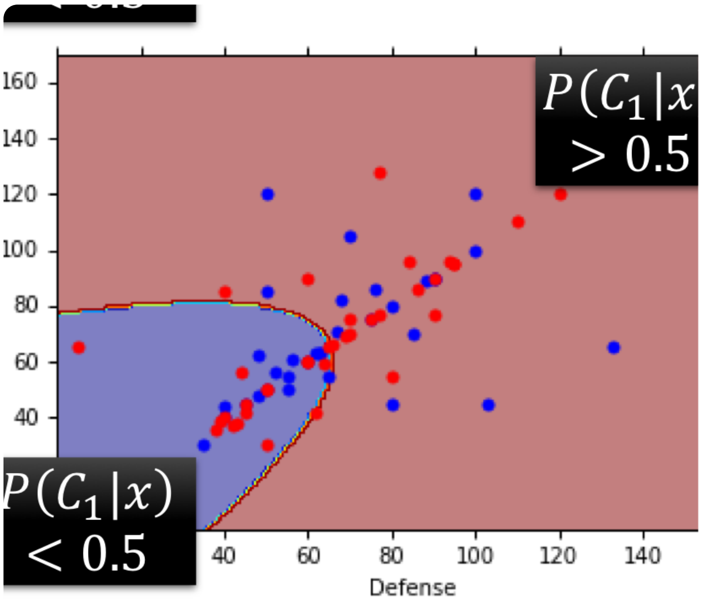
Σ^1, Σ^2: 7 x 7 matrices

54% accuracy … ☹



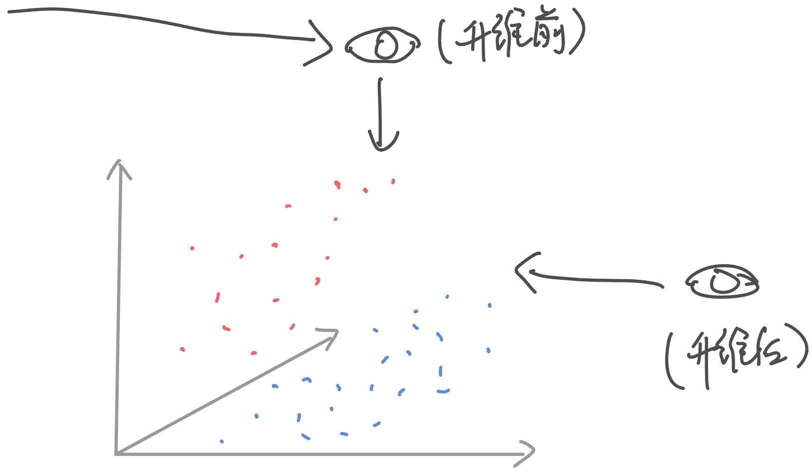
在二维feature训练结果准确率47%

升维度，

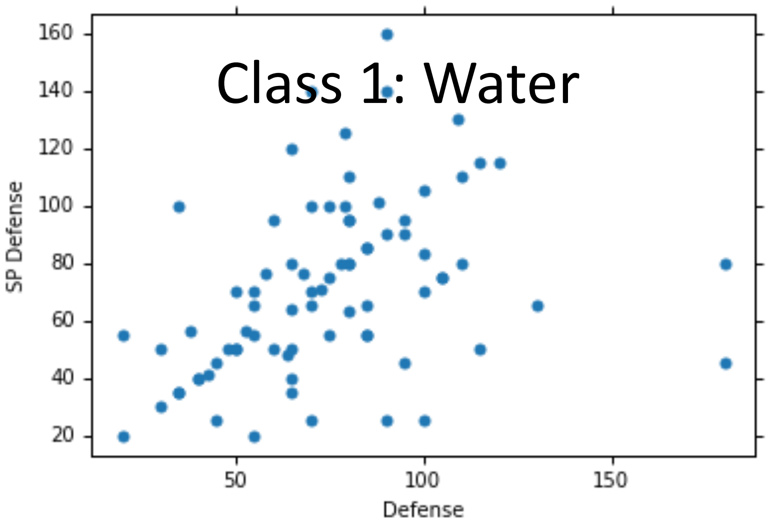


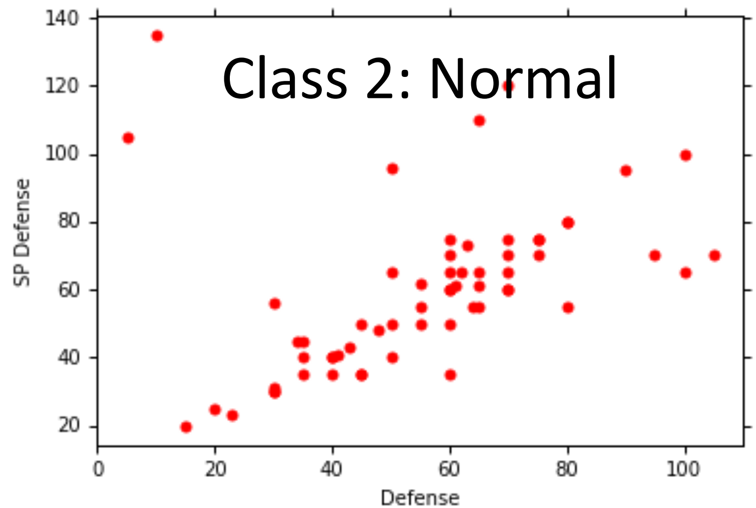
二维上看，蓝点和红点，糅在一起，
(从上往下看)

升维后，从三维空间看，点有不同的高度，
讲不定它们就分开了



Modifying Model





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

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

The same Σ

Less parameters

① 对每一类 C_i 确定对应的分布

② 将得到 $\mu^i \Sigma^i$

③ 现在 μ^1 与 $\mu^2 \dots \mu^i$ 共用同一个 Σ

$$\Sigma = P(\Sigma^1)\Sigma^1 + P(\Sigma^2)\Sigma^2 \dots P(\Sigma^i)\Sigma^i$$

④ $P(\Sigma^i) = \dfrac{C_i\text{样本数}}{\text{总样本数}}$

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 $\quad\Sigma\quad$ μ^2

共用同一个Σ

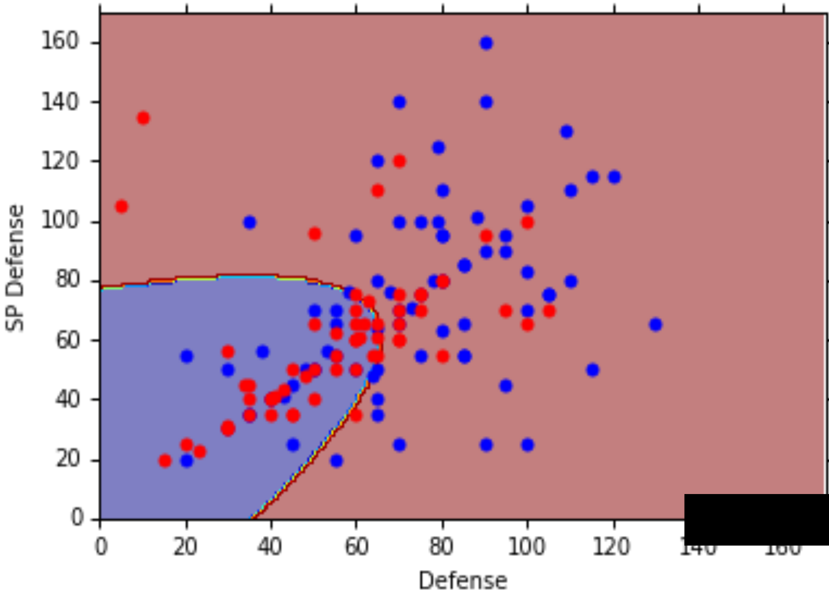
来减少计算量与参数量，避免overfitting

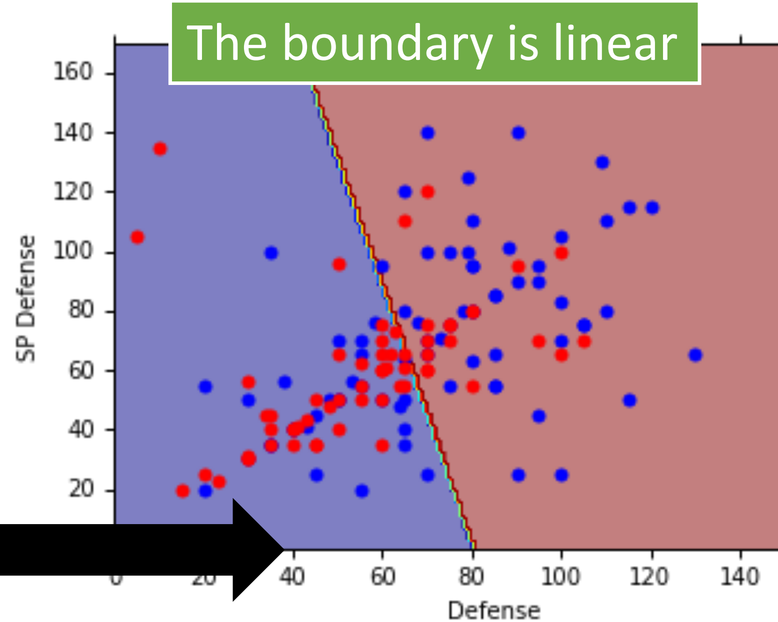
Find μ^1, μ^2, Σ maximizing the likelihood $L(\mu^1,\mu^2,\Sigma)$

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

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

Modifying Model





The same covariance matrix

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

54% accuracy → 73% accuracy

Three Steps

- Function Set (Model): 基本模型（贝叶斯分类）

$x \rightarrow$

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

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

Otherwise, output: class 2

- Goodness of a function: 找一组μ与Σ参数，即找到产生样本的高斯分布
  如何找：使产生训练数据的可能性最大的μ与Σ
  - 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 ☺

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

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

1-D Gaussian

Dimen

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*.

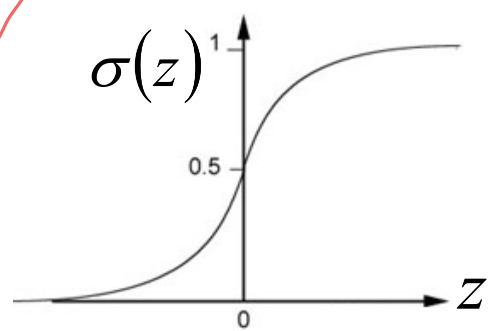
Posterior Probability

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

$$1 = \frac{1}{1 + \frac{P(x|C_2)P(C_2)}{P(x|C_1)P(C_1)}} \quad 3 = \frac{1}{1 + exp(-z)} \quad = \sigma(z)$$

Sigmoid function

设 $z = 2\ ln \frac{P(x|C_1)P(C_1)}{P(x|C_2)P(C_2)}$



Warning of Math

Posterior Probability

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\boxed{\begin{aligned} z = ln \frac{|\Sigma^2|^{1/2}}{|\Sigma^1|^{1/2}} &- \frac{1}{2}x^T(\Sigma^1)^{-1}x + (\mu^1)^T(\Sigma^1)^{-1}x - \frac{1}{2}(\mu^1)^T(\Sigma^1)^{-1}\mu^1 \\ &+ \frac{1}{2}x^T(\Sigma^2)^{-1}x - (\mu^2)^T(\Sigma^2)^{-1}x + \frac{1}{2}(\mu^2)^T(\Sigma^2)^{-1}\mu^2 + ln \frac{N_1}{N_2} \end{aligned}}$$

End of Warning

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

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

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

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

$\sigma(z) = \frac{1}{1+\exp(-z)} = \frac{1}{1+e^{-z}}$

$$P(C_1|x) = \sigma(\underbrace{w \cdot x + b}_{\text{线性函数}})$$

How about directly find *w* and b?

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

如果直接找w和b就变成了逻辑回归

Then we have ***w*** and b

总结

我们要求 $P(C_1|x)$ → 随机抽一个样本x，抽到x为 C_1 类的概率

① 假设所有样本服从于高斯分布
② 使用样本集确定所在高斯分布，训练，得到 μ, Σ，即密度函数 $f_{\mu,\Sigma}(x)$
③ 有了 $f_{\mu,\Sigma}(x)$，求得出了 $P(x|C_1)$ → 在某个分类中，抽到x的概率
④ 通过贝叶斯公式 $P(C_1|x) = \dfrac{P(x|C_1)P(C_1)}{P(x|C_1)P(C_1)+P(x|C_2)P(C_2)}$ #

这一章从样本服从高斯分布为前提，到训练高斯分布参数，再用这个分布去预测 $P(C_1|x)$

经整理(公式推导)发现，整个逻辑可简化到如下形式

$$P(C_1|x) = \sigma(\underline{w\cdot x + b}) = \frac{1}{1+\exp(-z)}$$

我们熟悉的线性模型

所以，其实这一章是基于高斯分布定义了一个复杂线性模型

$$P(C_1|x) = 贝叶斯(f_{\mu,\Sigma}(x), \ldots) \Rightarrow \underline{\sigma(w\cdot x + b)}$$

我们也可以直接定义线性模型 $(wx+b)$
就变成了简单的线性回归

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