Classification:Probabilistic Generative Model

Classification

x → Function → Class n

- Credit Scoring
  - Input: income, savin
    history ......
  - Output: accept or re
- Medical Diagnosis
  - Input: current symp
    history ......
  - Output: which kind

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?

How to do Classification

- Training data for Classification

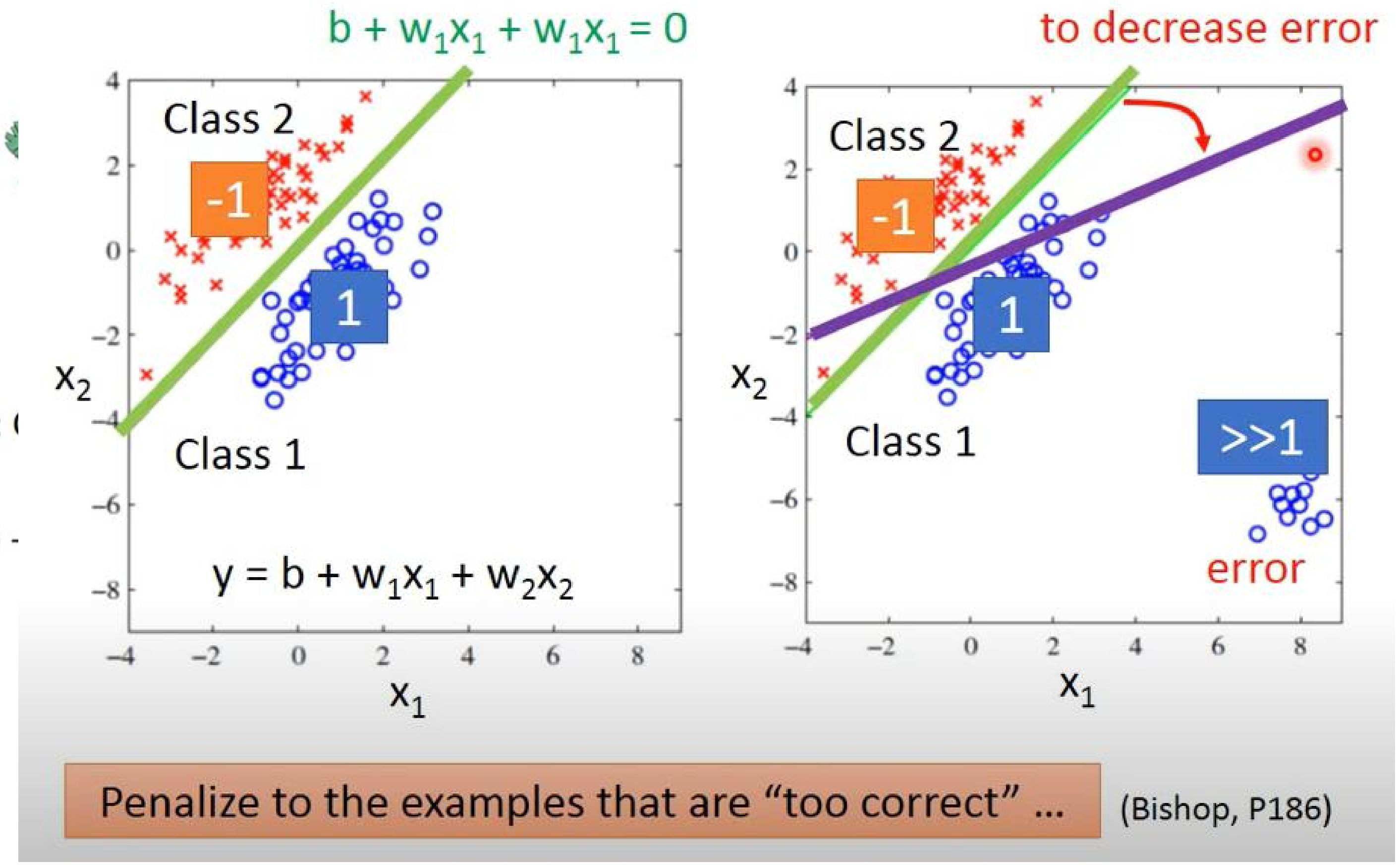
$(x^1, \hat{y}^1)$ $(x^2, \hat{y}^2)$ ……

Classification as Regression?

Binary classification as example

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

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



注：该图为三维图像在二维图像上的投影，颜色表示y的大小

Ideal Alternatives

- Function (Model):

x

- Loss function:

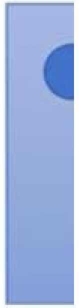
$$L(f) = \sum_n \delta(f(x$$

- Find the best funct
  - Example: Perce

Two Cla

Class 1

$P(C_1)$

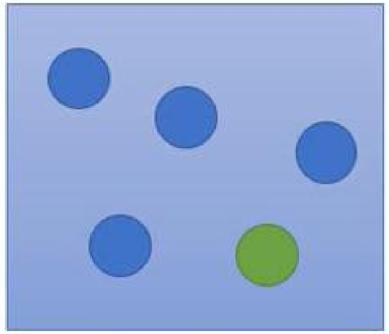


Given an x, v

Prior

Class 1

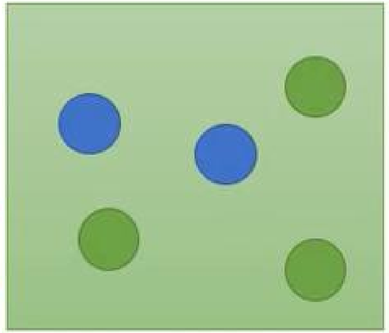
$P(C_1)$



Water

Class 2

$P(C_2)$



Normal

Water and Normal type with ID < 400 for training,
rest for testing

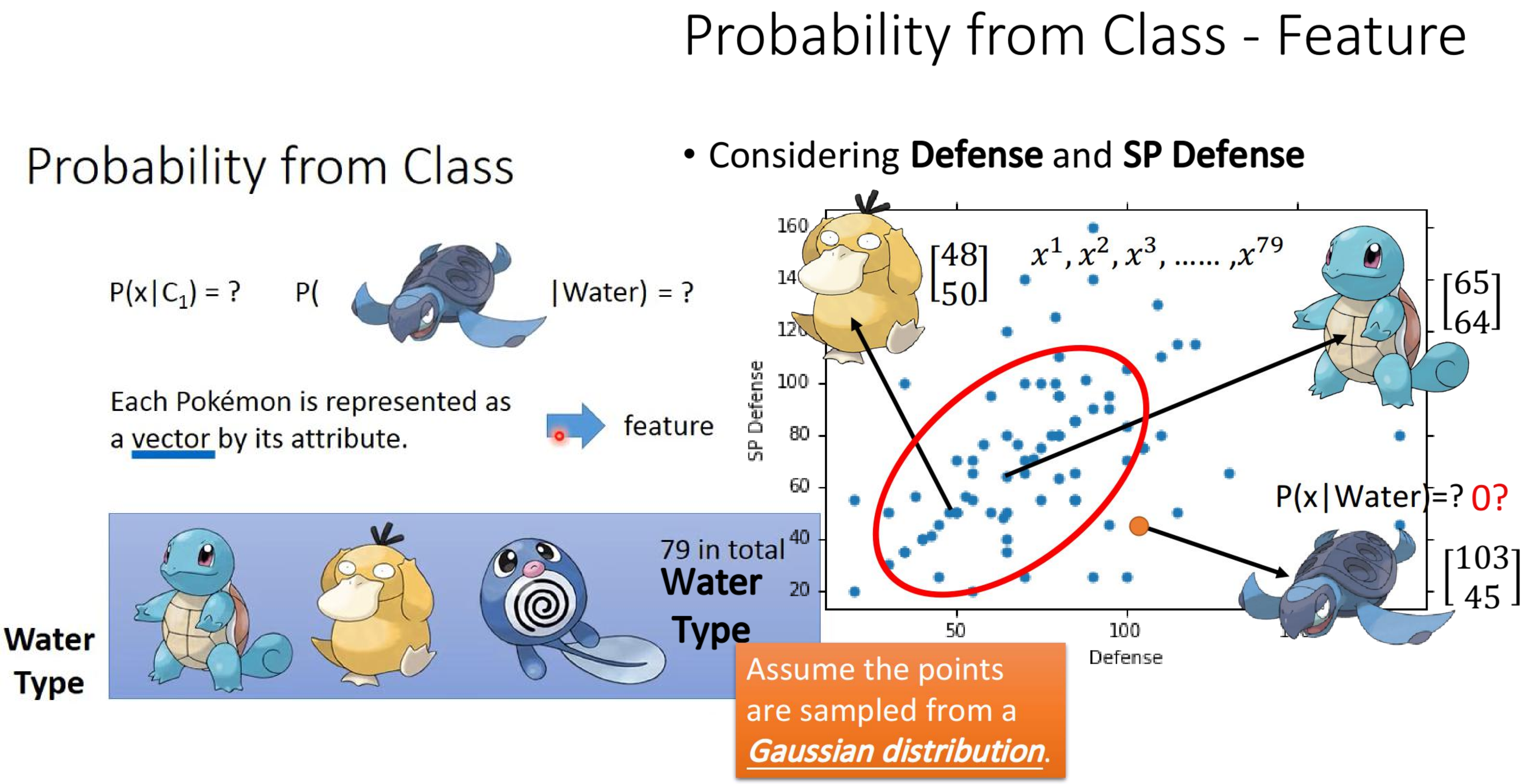
Training: 79 Water, 61 Normal

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

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

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

P(x|c1)

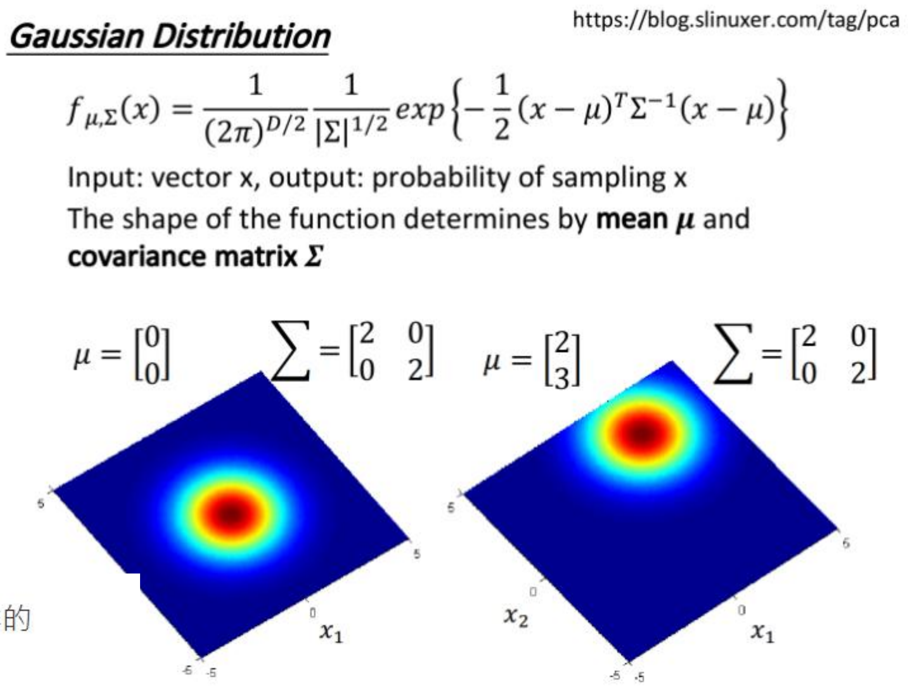

Probability from Class - Feature
Probability from Class
P(x|C1) = ?
P(|Water) = ?
Each Pokémon is represented as a vector by its attribute.
feature
• Considering Defense and SP Defense
$x^1, x^2, x^3, \ldots\ldots, x^{79}$
$\begin{bmatrix}48\\50\end{bmatrix}$
$\begin{bmatrix}65\\64\end{bmatrix}$
$\begin{bmatrix}103\\45\end{bmatrix}$
P(x|Water)=? 0?
SP Defense
Defense
79 in total
Water Type
Water Type
Assume the points are sampled from a Gaussian distribution.

Gaussian Distribution

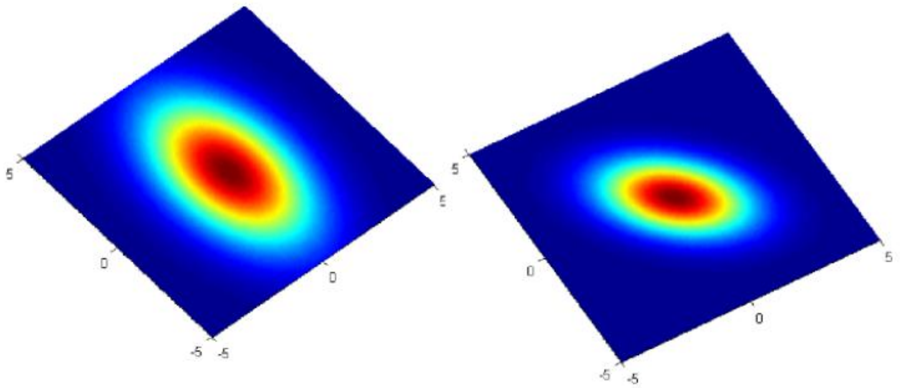
先介绍一下高斯函数，这里u表示均值，Σ表示方差，两者都是矩阵matrix，那高斯函数的概率密度函数则是：

$$f_{u,\Sigma}(x) = \frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{|\Sigma|^{\frac{1}{2}}}e^{-\frac{1}{2}(x-u)^T\Sigma^{-1}(x-u)}$$

从下图中可以看出，同样的Σ，不同的u，概率分布最高点的地方是不一样的



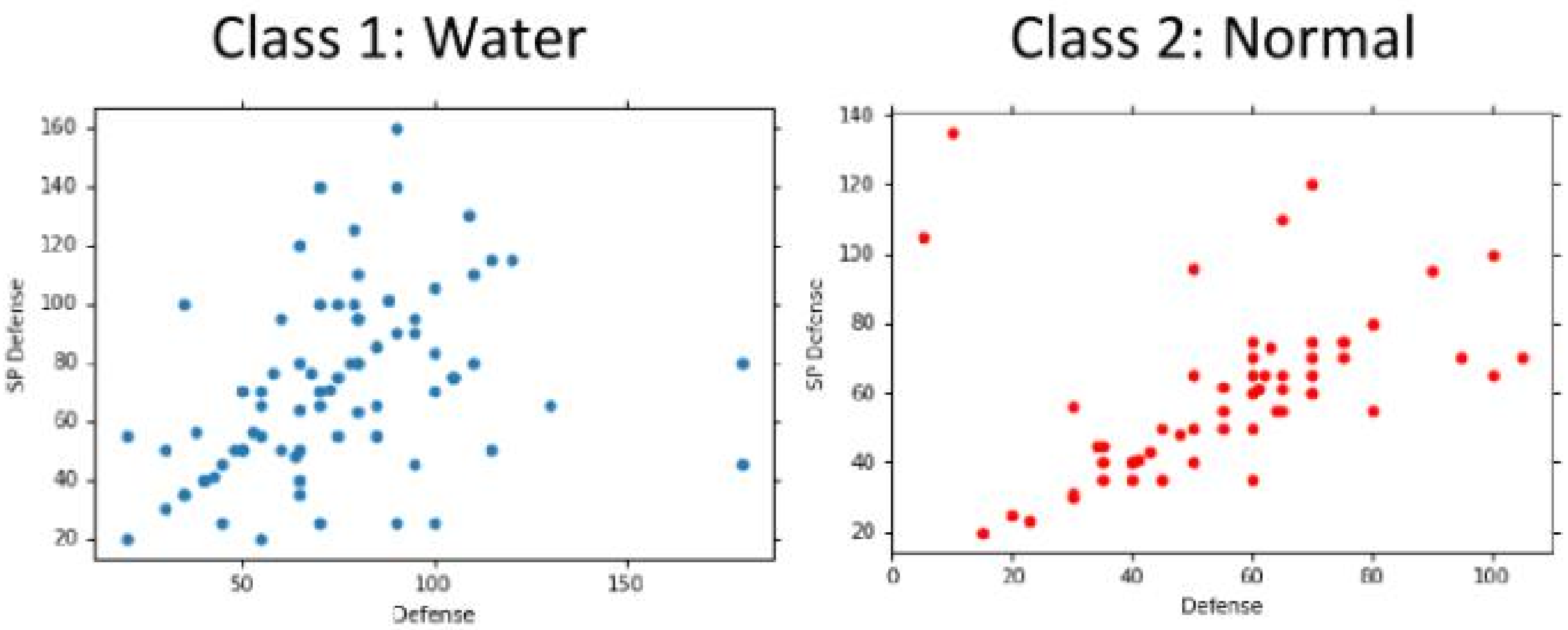
同理，如果是同样的u，不同的Σ，概率分布最高点的地方是一样的，但是分布的密集程度是不一样的



Maximum Likelihood

极大似然估计：找出最特殊的那对u和Σ，从它们共同决定的高斯函数中再次采样出79个点，使"得到的分布情况与当前已知79点的分布情况相同"这件事发生的可能性最大

Maximum Likelihood



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

Do Classfication

Now we can d

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

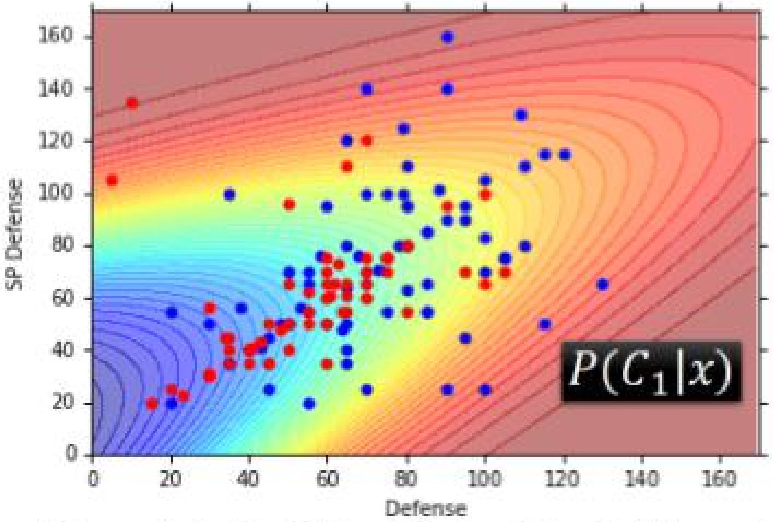
$$\mu^1 = \begin{bmatrix} 75.0 \\ 71.3 \end{bmatrix} \quad \Sigma^1 = \begin{bmatrix} 874 & 327 \\ 327 & 929 \end{bmatrix}$$

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

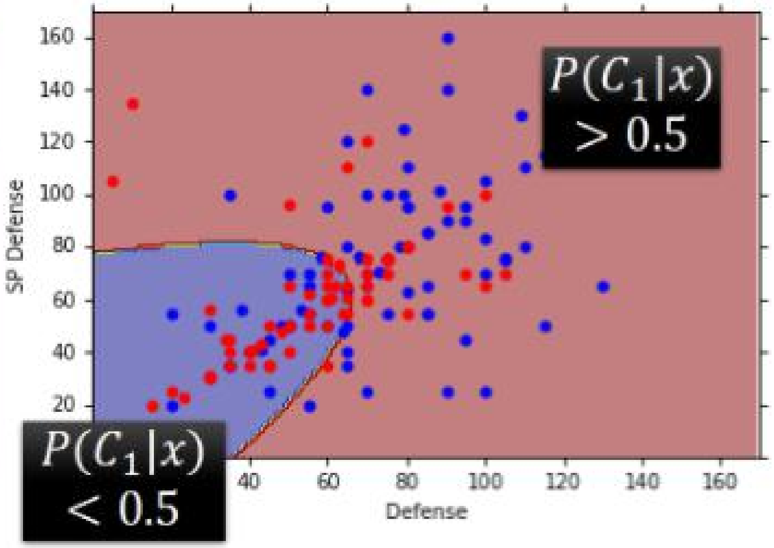
$$f_{\mu^2,\Sigma^2}(x) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma^2|^{1/2}} exp\left\{-\frac{1}{2}\right.$$

$$\mu^2 = \begin{bmatrix} 55.6 \\ 59.8 \end{bmatrix} \quad \Sigma^2 = \begin{bmatrix} 847 & 42 \\ 422 & 68 \end{bmatrix}$$

If $P(C_1|x) > 0.5$



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



How's the results?

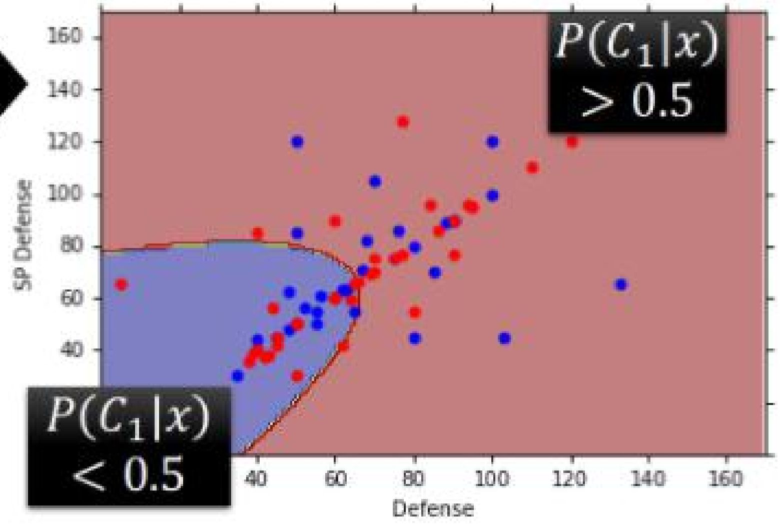
Testing data: 47% accuracy ☹

All: hp, att, sp att, de, sp de, speed (6 features)

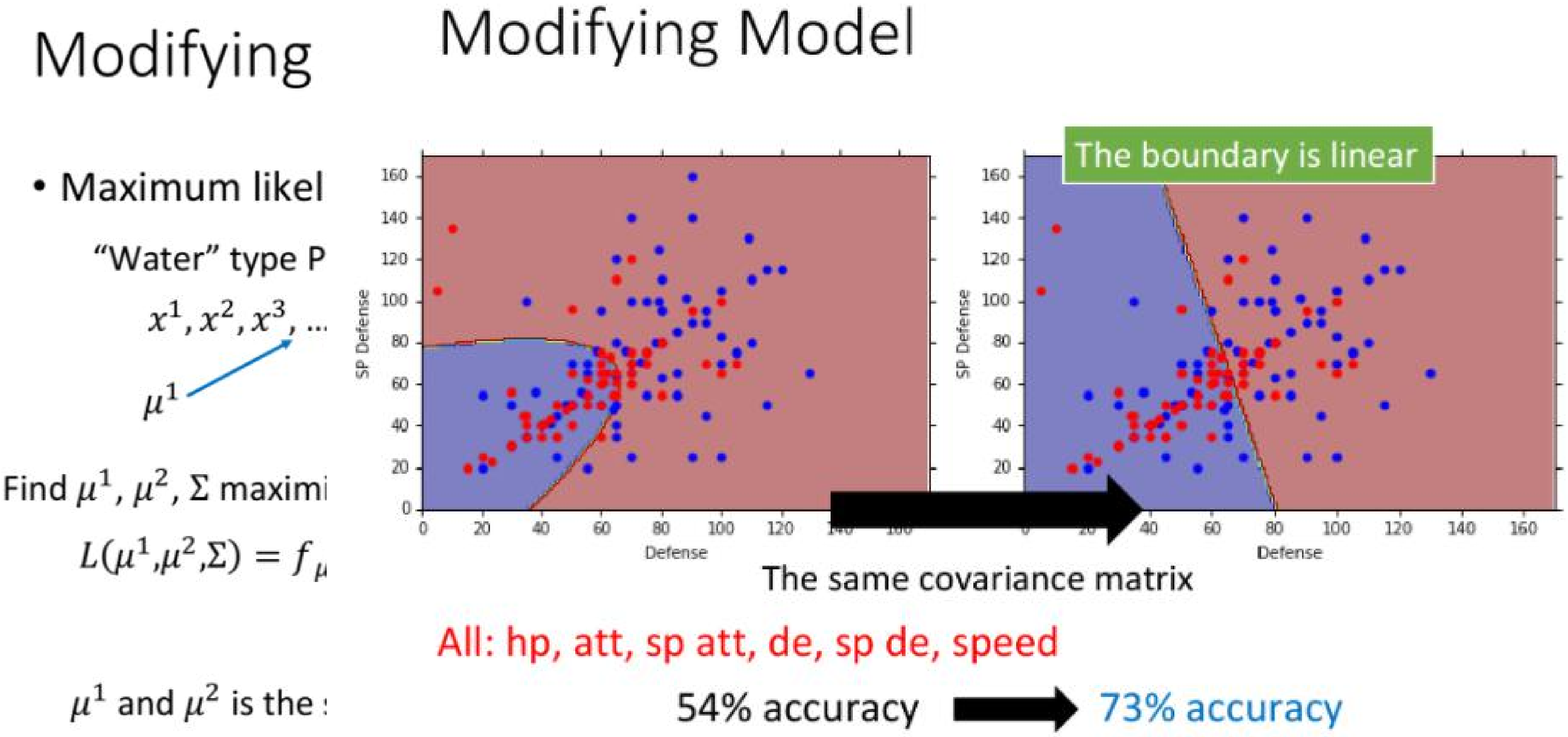
μ^1, μ^2: 6-dim vector

Σ^1, Σ^2: 6 x 6 matrices

64% accuracy …



Modifying Model



Three Steps

- Function Set (Model):

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

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

Otherwise, output: class 2

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