

# 李宏毅机器学习-P9

## Classification

### Classification

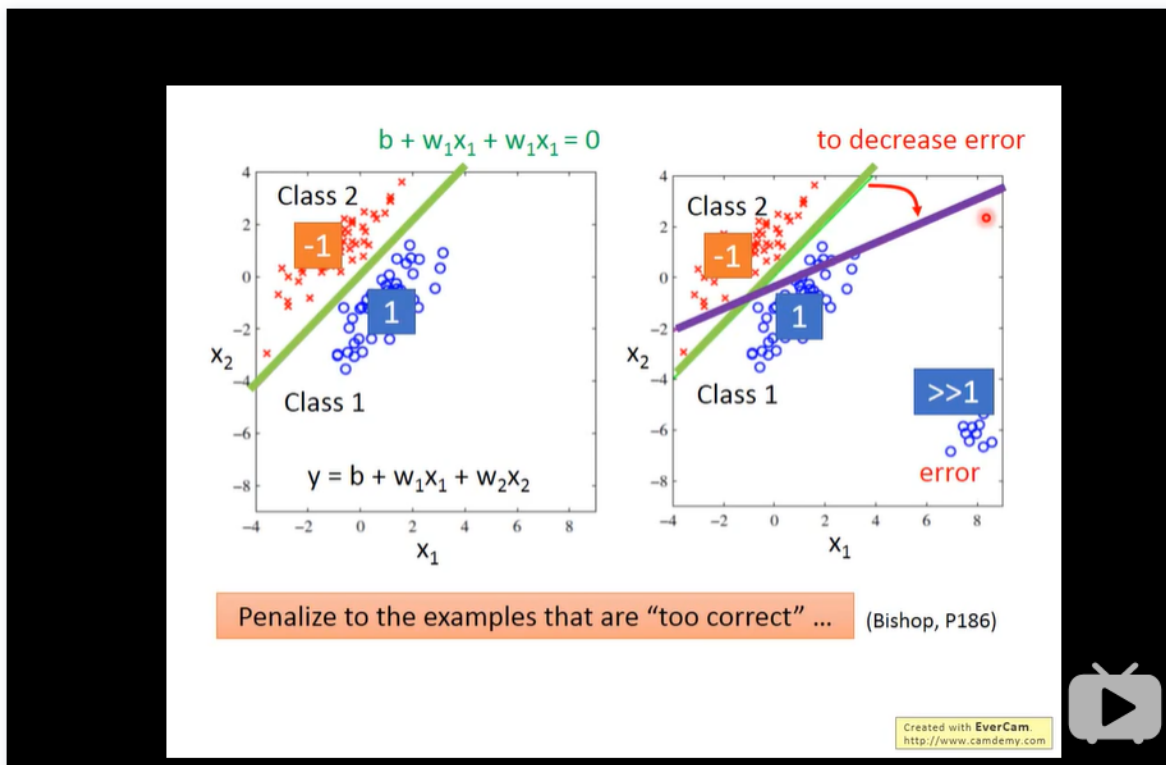
- Credit Scoring
  - Input: income,savings,profession,age, past financial history...
  - Output: accept or refuse
- Medical Diagnosis
  - Input: current symptoms,age,gender,past medical history...
  - Output: which kind of diseases
- Handwritten character recognition
- Face Recognition

### How to Classification

#### Classification as Regression?

Binary(二元的) classification as example

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



- 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 Alternative

没办法微分

理想替代

• Function (Model):  $f(x)$

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

• 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

Created with EverCam  
<http://www.camdemy.com>

## Two Boxes

以从盒子里拿两种不同颜色的球为例，计算概率

(贝叶斯公式)

(全概率公式)

## Probability from Class

以求从水系神奇宝贝中抽一只海龟的几率为例

每个神奇宝贝都以一个向量来表示，向量储存着其信息

但问题是海龟不一定在training data里面出现，但我们不可能说概率是零。

那么我们这时候 assume the points are sampled from a Gaussian distribution (高斯分布)

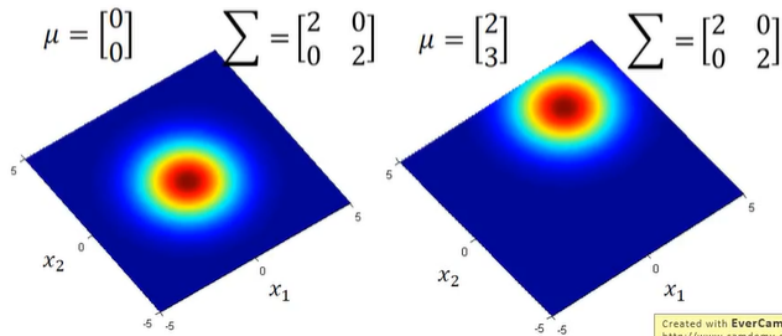
### Gaussian Distribution

<https://blog.slinuxer.com/tag/pca>

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

Input: vector  $x$ , output: probability of sampling  $x$

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



对于一个新的点 $x$ ,我们通过计算就可以知道他被sample出来的几率有多大（记得去看高斯分布的计算方法）越接近中心点 $\mu$ 被sample的几率越大。那么要找到想要的 $\mu$ 和 $\Sigma$ ，这里用的概念叫Maximum Likelihood

### Maximum Likelihood

The Gaussian with any mean  $\mu$  and covariance matrix  $\Sigma$  can generate these points(Different Likelihood)

Likelihood of a Gaussian with mean  $\mu$  and covariance matrix  $\Sigma$ 、 = the probability of the Gaussian samples  $x^1, x^2, x^3, \dots, x^n$

# Maximum Likelihood

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

We assume  $x^1, x^2, x^3, \dots, x^{79}$  generate from the Gaussian  $(\mu^*, \Sigma^*)$  with the **maximum likelihood**

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

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

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

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

average

Created with EverCam.  
<http://www.camdemy.com>

故意使对象共用 covariance matrix  $\Sigma$  就可以使用 less parameter

## Modifying Model

Ref: Bishop,  
chapter 4.2.2

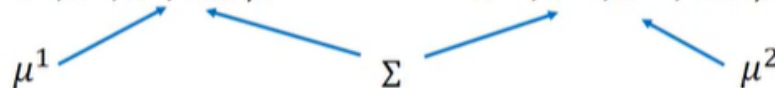
### • Maximum likelihood

"Water" type Pokémons:

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

"Normal" type Pokémons:

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



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

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

$\mu^1$  and  $\mu^2$  is the same

$$\Sigma = \frac{79}{140} \Sigma^1 + \frac{61}{140} \Sigma^2$$

Created with EverCam.  
<http://www.camdemy.com>

当共用 covariance matrix  $\Sigma$  时 boundary 就会变成 linear 的

### Three Steps

- Function Set(Model)

$$P(c_1 | x) = \frac{p(x|c_1)p(c_1)}{p(x|c_1)p(c_1) + p(x|c_2)p(c_2)}$$

If  $p(c_1 | x) > 0.5$ , output: class 1

otherwise, output: class 2

- Goodness of a function:

- The mean  $\mu$  and covariance  $\Sigma$  that maximizing the likelihood (the probability of generating data)

- Find the best function: easy

## Probability Distribution

- You can always use the distribution you like

If you assume all the dimensions are independent then you are using Naive Bayes Classifier

## Posterior Probability

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

$\sigma(z)$  --Sigmoid function

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

## Warning of Math(变戏法)

### Posterior Probability

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

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

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

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

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

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

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

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

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

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

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

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

$$P(C_1|x) = \sigma(\mathbf{w} \cdot x + b) \quad \text{How about directly find } \mathbf{w} \text{ and } b?$$

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

Then we have  $\mathbf{w}$  and  $b$

Created with EverCam.  
http://www.camdemy.com

$$\begin{aligned} z &= \ln \frac{P(x|C_1)}{P(x|C_2)} + \ln \frac{P(C_1)}{P(C_2)} = \frac{N_1}{N_2} \\ &= \ln \frac{|\Sigma^2|^{1/2}}{|\Sigma^1|^{1/2}} - \frac{1}{2} [(x - \mu^1)^T (\Sigma^1)^{-1} (x - \mu^1) - (x - \mu^2)^T (\Sigma^2)^{-1} (x - \mu^2)] \\ &\quad (x - \mu^1)^T (\Sigma^1)^{-1} (x - \mu^1) \\ &= x^T (\Sigma^1)^{-1} x - \underline{x^T (\Sigma^1)^{-1} \mu^1 - (\mu^1)^T (\Sigma^1)^{-1} x} + (\mu^1)^T (\Sigma^1)^{-1} \mu^1 \\ &= x^T (\Sigma^1)^{-1} x - \underline{2(\mu^1)^T (\Sigma^1)^{-1} x} + (\mu^1)^T (\Sigma^1)^{-1} \mu^1 \\ &\quad (x - \mu^2)^T (\Sigma^2)^{-1} (x - \mu^2) \\ &= x^T (\Sigma^2)^{-1} x - 2(\mu^2)^T (\Sigma^2)^{-1} x + (\mu^2)^T (\Sigma^2)^{-1} \mu^2 \end{aligned}$$

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

Created with EverCam.  
http://www.camdemy.com

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

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

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

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

$$P(C_1|x) = \sigma(w \cdot x + b) \quad \text{How about directly find } w \text{ and } b?$$

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

Then we have  $w$  and  $b$

Created with EverCam.  
<http://www.camdemy.com>

经过计算我们不难理解为什么当  $\Sigma_1 = \Sigma_2$  的时候边界会变成linear的。

但是我们能不能直接找到  $w$  和  $b$  呢？