李宏毅机器学习-P9

Classification

Classification

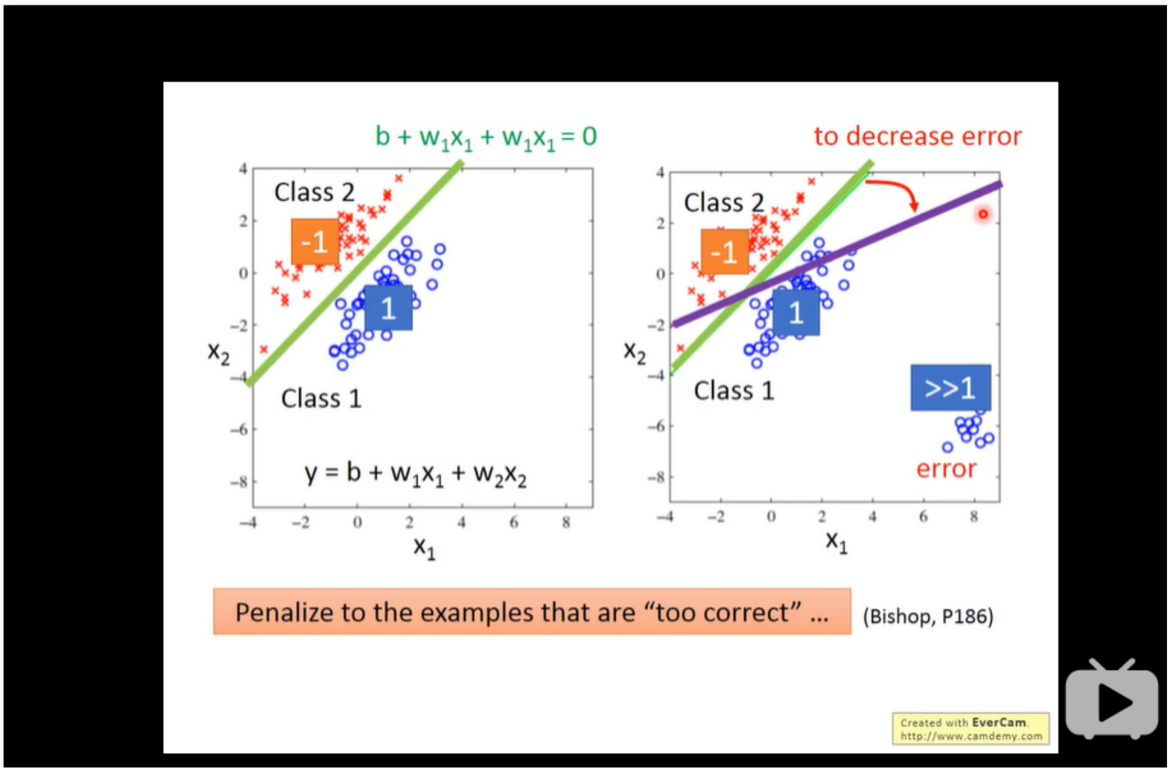
- Credit Scoring
  - Input: income,savings,profession,age, past financial history· · ·
  - Output: accept or refuse
- Medical Diagnosis
  - Input: current symptoms,age,gender,past medical history· · ·
  - Output: which kind of diseases
- Handwritten character recognition
- Face Recognition

How to Classification

Classification as Regression?

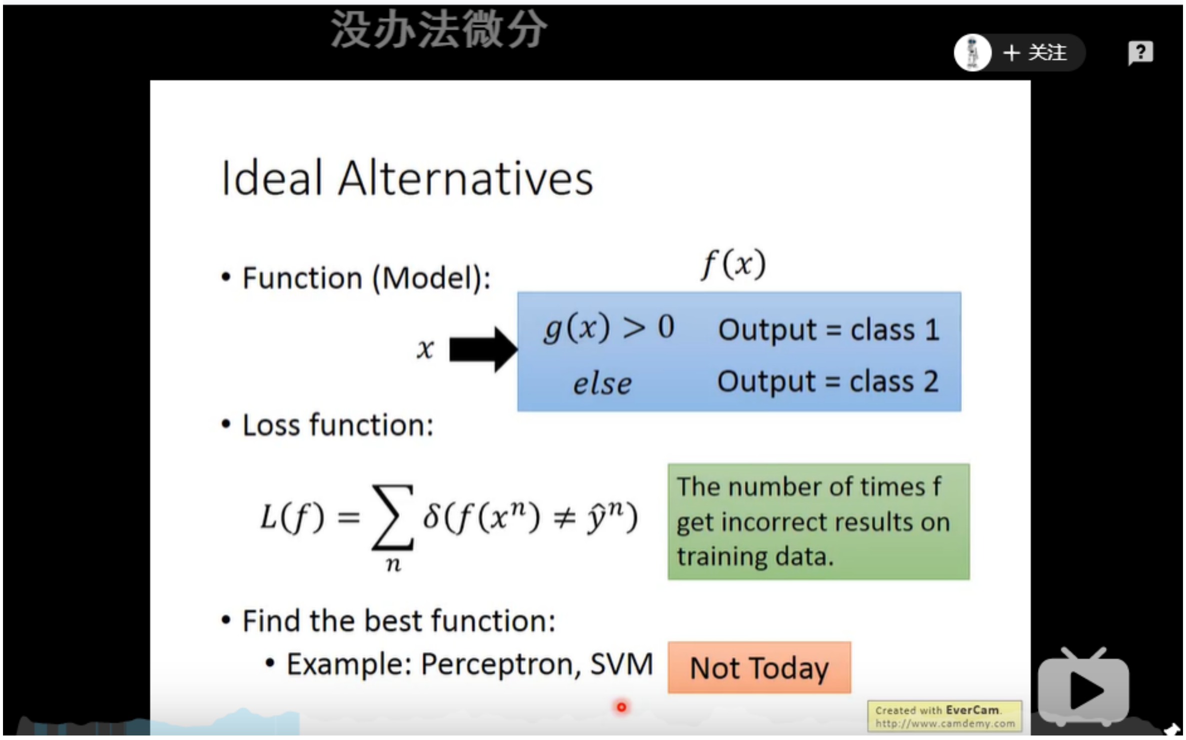
Binary(二元的) classification as example

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



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）

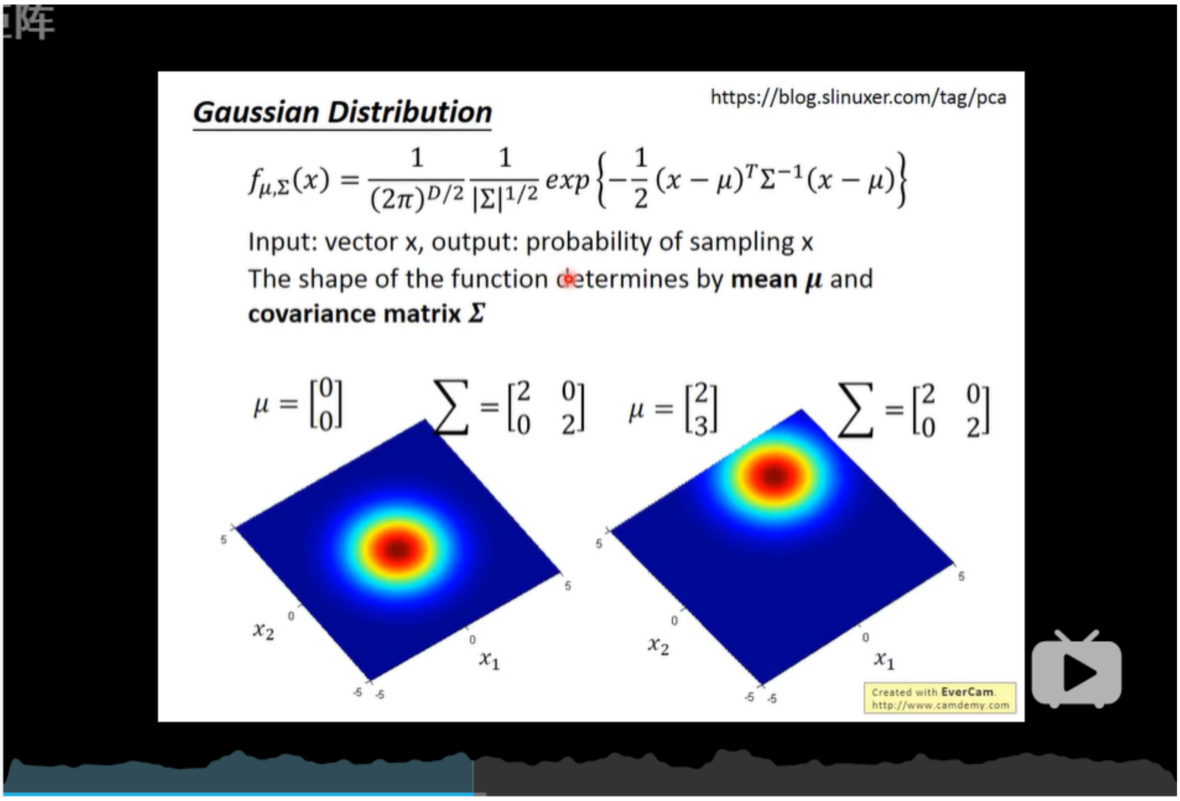
Probability from Class

以求从水系神奇宝贝中抽一只是海龟的几率为例

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）

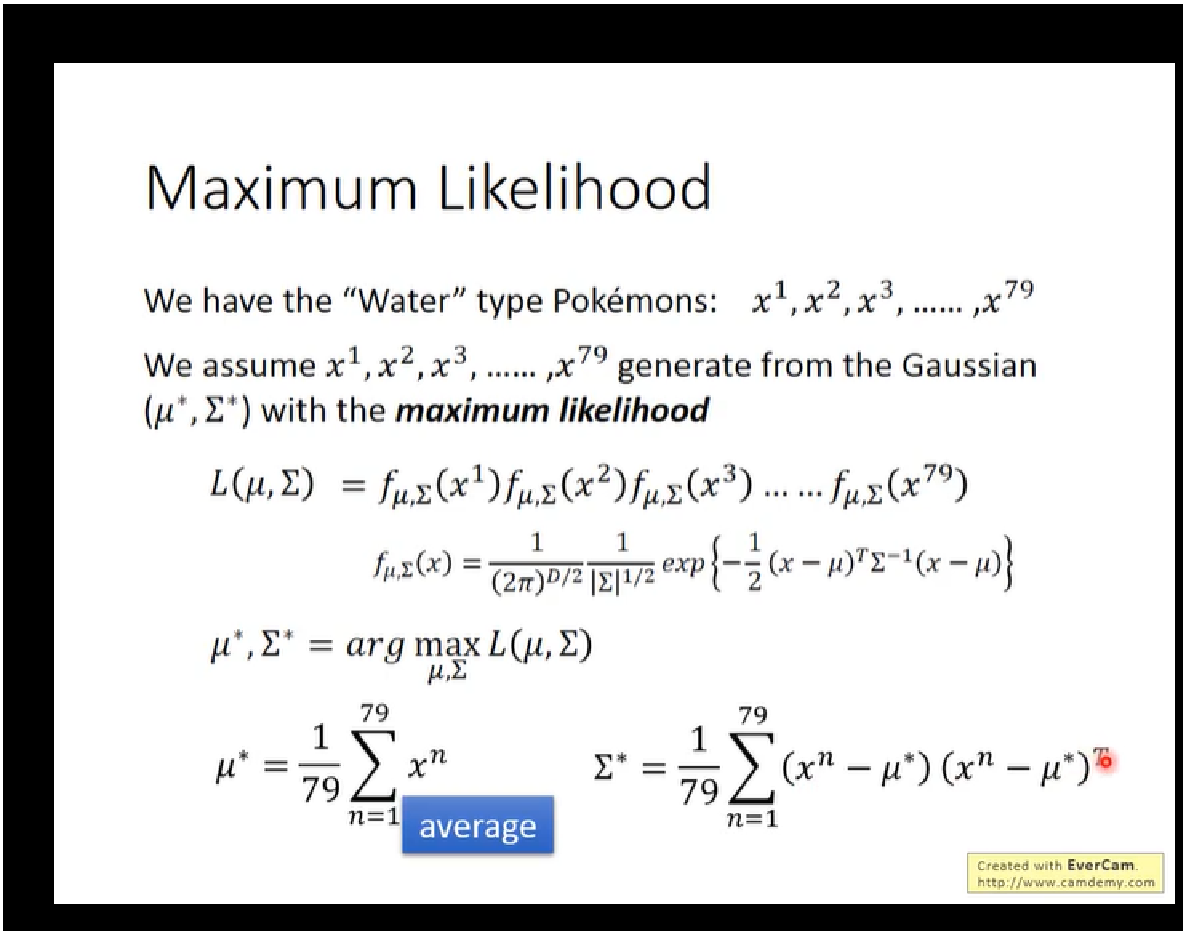


对于一个新的点x,我们通过计算就可以知道他被sample出来的几率有多大（记得去看高斯分布的计算方法）越接近中心点μ被sample的几率越大。那么要找到想要的μ和Σ，这里用的概念叫Maximum Likelihood

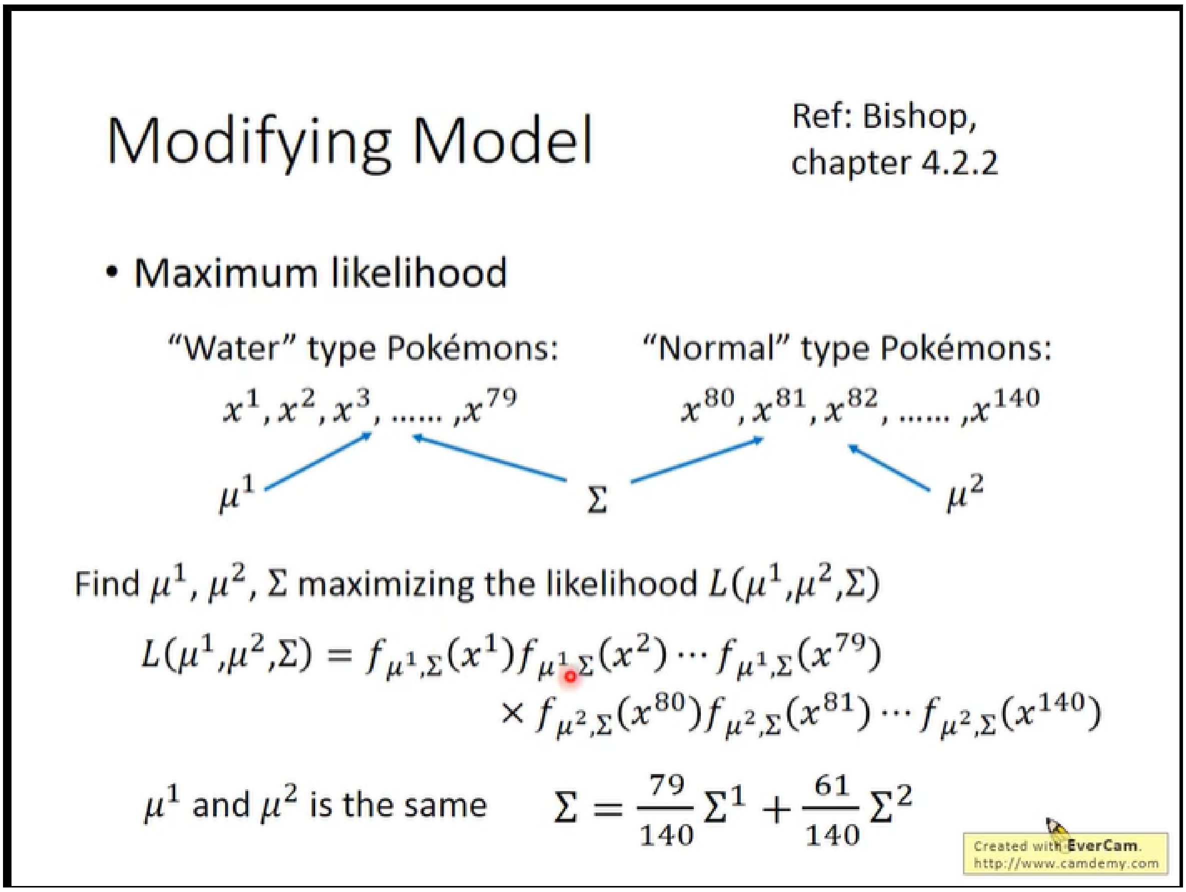
Maximum Likelihood

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 ,$\cdots$,x^n



故意使对象共用 covariance matrix Σ 就可以使用 less parameter



当共用 covariance matrix Σ 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 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

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

Posterior Probability

$P(C_1|x)$ = $\frac{P(x|C_1)P(C_1)}{P(x|C_1)P(C_1)+P(x|C_2)P(C_2)}$ =$\frac{1}{1+\frac{p(x|C_3)P(C_2)}{P(x|C_1)P(C_1)}}$ =$\frac{1}{1+exp(-z)}$ =σ (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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

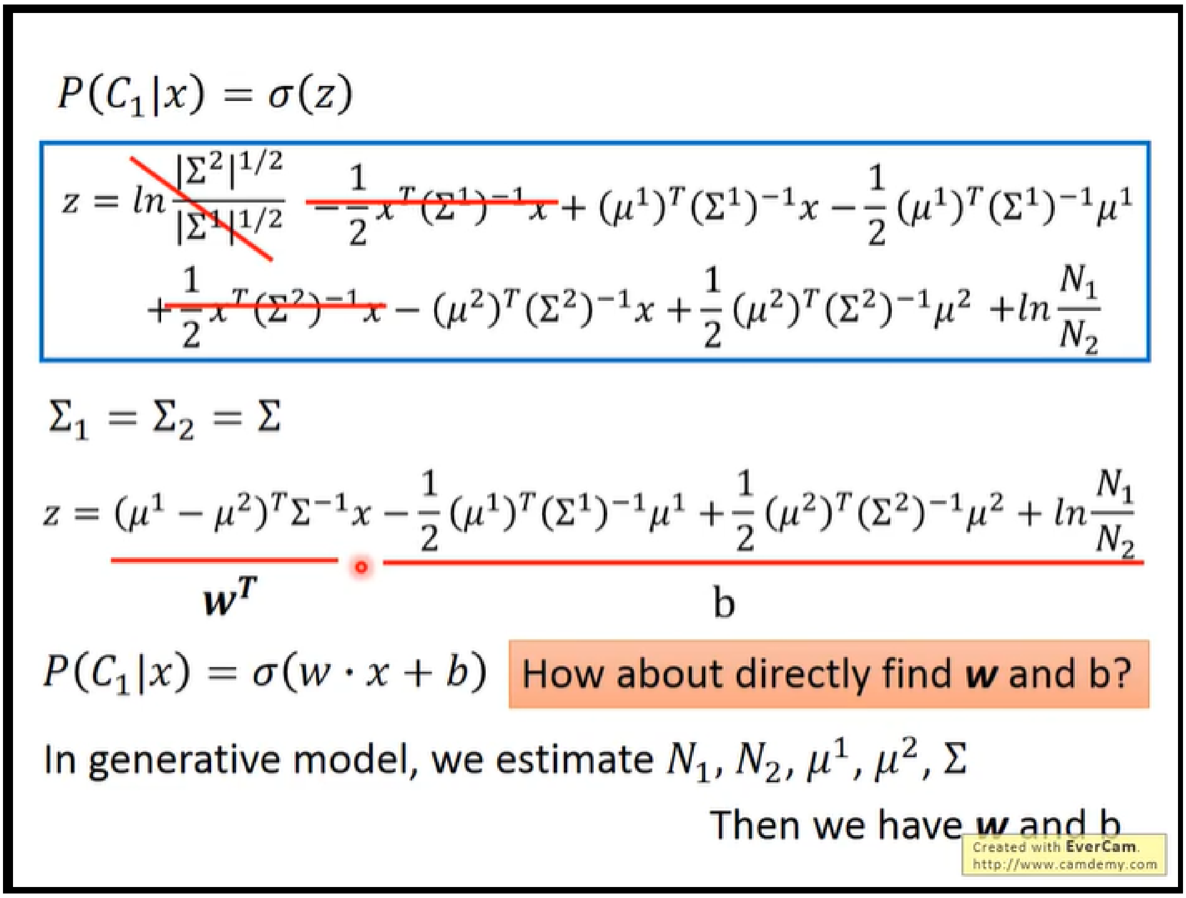
$$= x^T(\Sigma^1)^{-1}x - 2(\mu^1)^T(\Sigma^1)^{-1}x + (\mu^1)^T(\Sigma^1)^{-1}\mu^1$$

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

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

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

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?