... should be answered in **ENGLISH on the answer sheets**. This is an open-book examination so that you can use any materials at hand (except laptops and pads).

1. Suppose we have two categories ω_1 and ω_2, and the training examples (one dimensional) for each category are given as follows:

training example

$$D_1 = \{-6, -4, -3, -1, 0, 2\} \quad -2$$

$$D_2 = \{-2, 1, 4, 5\} \quad 2$$

Please answer the following questions:

a) The *Bayes Theorem* is named after Thomas Bayes (1702-1761). What is his nationality and occupation? (6 points)

England

~~London~~ statistician, philosopher, $\boxed{\text{priest}}$ minister and presbyterian -0.5 -2

b) Assume that the likelihood function of each category has normal parametric form, i.e. $p(x \mid \omega_1) \sim N(\mu_1, \sigma_1^2)$ and $p(x \mid \omega_2) \sim N(\mu_2, \sigma_2^2)$. Which category should we decide on $x = -0.5$ when *maximum-likelihood estimation* is employed to make the prediction? (**10 points**) $-0.5+2$

PPT 3-11~13 PPT 3-3

c) Based on the maximum-likelihood estimates for μ_i and σ_i^2 $(i = 1, 2)$, suppose we further have a set of three possible actions $A = \{\alpha_1, \alpha_2, \alpha_3\}$ with the following loss function: PPT 2-25~29

category \ *action*	$\alpha = \alpha_1$	$\alpha = \alpha_2$	$\alpha = \alpha_3$
$\omega = \omega_1$	5 λ_{11}	10	20
$\omega = \omega_2$	10	30	3

$$\frac{9+1+16}{26}$$

For $x = -0.5$, which action should we apply on it? **(8 points)**

d) Following the assumptions in b), suppose we further know that $\sigma_1 = 1$, $\sigma_2 = 2$, $\mu_1 \sim N(-2, 1)$ and $\mu_2 \sim N(2, 1)$. Which category should we decide on $x = -0.5$ when *Bayesian estimation* is employed to make the prediction? **(8 points)** PPT 3-19

e) Assume that the likelihood function of each category doesn't have any parametric form. Which category should we decide on $x = -0.5$ when *k_n-nearest neighbor* is employed to make the prediction? Here, k_n=2 and the cell (i.e. *line segment*) grows around x uniformly. **(8 points)** PPT4-19

2. Given a set of three states of nature $\Omega = \{\omega_1, \omega_2, \omega_3\}$ with prior probabilities $P(\omega_1) = 0.4$, $P(\omega_2) = 0.4$ and $P(\omega_3) = 0.2$. Furthermore, each pattern is represented by a two-dimensional feature vector $\mathbf{x}$ $=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ with Gaussian class-conditional pdf $p(\mathbf{x} \mid \omega_i) \sim N(\mu_i, \Sigma_i)$ $(1 \le i \le 3)$. Specifically, we have mean vectors $\mu_1 = (0, 0)^t$, $\mu_2 = (-2, -1)^t$, $\mu_3 = (1, 2)^t$ and covariance matrix $\Sigma_1 = \Sigma_2 = \Sigma_3 = \begin{pmatrix} 3 & 3 \\ 3 & 6 \end{pmatrix}$.

Please answer the following questions:

a) Prove that $\Sigma_i^{-1} = \begin{pmatrix} 2/3 & -1/3 \\ -1/3 & 1/3 \end{pmatrix}$ $(1 \le i \le 3)$; **(5 points)**

正定矩阵

b) Prove that Σ_i $(1 \le i \le 3)$ is *positive definite*; **(5 points)** PPT2-43/45

c) Specify the discriminant function for each state of nature; **(6 points)**

选择 PPT2-46 一致 / 分Case

d) Given a pattern $\mathbf{x} = (-2, -2)^t$, which state of nature should we decide on based on the above discriminant functions? **(4 points)** 选 $g_j(x)$ 大的

3. Let x have a *Rayleigh density* $(\theta > 0)$

$$p(x \mid \theta) = \begin{cases} 2\theta x\, e^{-\theta x^2} & x \ge 0 \\ 0 & \text{otherwise} \end{cases}$$ $\left(-1, \frac{1}{3}\right)$

a) Show that $p(x \mid \theta)$ is indeed a pdf function; **(10 points)**

b) Suppose that n examples $\{x_1, \ldots, x_n\}$ are drawn independently according to $p(x \mid \theta)$. Show that the *maximum-likelihood estimate* for θ is given by

$$\hat{\theta} = 1/\left(\frac{1}{n}\sum_{k=1}^{n} x_k^2\right)$$ **(15 points)**

4. Suppose we have four training examples under the two-category case, i.e. $\mathcal{D}^* = \{(\mathbf{x}_i, \omega_i) \mid 1 \leq i \leq 4\}$ where $\mathbf{x}_1 = (1,1)'$, $\mathbf{x}_2 = (2,1)'$, $\mathbf{x}_3 = (-1,-2)'$, $\mathbf{x}_4 = (2,-3)'$ and $\omega_1 = \omega_2 = -1$, $\omega_3 = \omega_4 = +1$. Furthermore, linear discriminant function $g(\mathbf{x}) = \mathbf{w}'\mathbf{x} + b$ is adopted to learn from the training examples and the criterion function to be minimized is set as $J(\mathbf{w}, b) = \frac{1}{2}\mathbf{w}'\mathbf{w} - \sum_{i=1}^{4} \omega_i \cdot g(\mathbf{x}_i)$.

Given the initial model $\mathbf{w}_0 = (2,1)^t$ and $b_0 = -1$. Please answer the following questions:

$J(w, b)$

a) If *gradient descent* techniques are utilized to minimize the criterion function, what is the resulting discriminant function after *three* gradient descent steps with learning rate $\eta = 0.1$ and threshold $\epsilon = 10^{-6}$? **(12 points)** PPT5-10

b) Given a pattern $\mathbf{x} = (-2,-2)'$, should we decide *positive* (+1) or *negative* (−1) on $\mathbf{x}$? **(3 points)** PPT5-6

(1.4 , -1.1) PPT5-6

Final Exam

1. a) ✓

b) For $p(x|w_1) \sim N(\mu_1, \sigma_1^2)$, can know (PPT 3-11) μ, and Σ unknow situation

$$p(x_k|\theta) = \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{\left[-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right]} \quad \left(\theta = \begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix} = \begin{bmatrix}\mu_1\\ \sigma_1^2\end{bmatrix}\right) \star$$

apply maximum-likelihood estimation.

$$\nabla_\theta \ln p(x_k|\theta) = \begin{bmatrix} \frac{1}{\theta_2}(x_k - \theta_1) \\ -\frac{1}{2\theta_2} + \frac{(x_k-\theta_1)^2}{2\theta_2^2} \end{bmatrix}$$

let $\nabla_\theta \ln p(x_k|\theta)\big|_{\theta=\hat\theta} = 0$

then $\hat\mu_1 = \frac{1}{n}\sum_{k=1}^{n} x_k$, $\hat\sigma_1^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \hat\mu_1)^2$ ① (x_k from D_1)

same as ① caculate, can get:

$\hat\mu_2 = \frac{1}{n}\sum_{k=1}^{n} x_k$, $\hat\sigma_2^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \hat\mu_2)^2$ ② (x_k from D_2)

$\because D_1 = \{-6, -4, -3, -1, 0, 2\} \rightarrow p_1(x|w_1) \sim N(-2, 7)$ $\left\{\hat\mu_1 = -2, \hat\sigma_1^2 = 7\right.$

$D_2 = \{-2, 1, 4, 5\} \rightarrow p_2(x|w_2) \sim N(2, \frac{15}{2})$ $\left\{\hat\mu_2 = 2, \hat\sigma_2^2 = \frac{15}{2}\right.$

$\therefore$ (when $x=-0.5$) according to $\star$

$$\begin{cases} p(x|w_1) = \frac{1}{\sqrt{2\pi\cdot 7}} e^{\left[-\frac{(-0.5+2)^2}{2\times 7}\right]} = 0.128 \\ p(x|w_2) = \frac{1}{\sqrt{2\pi\cdot\frac{15}{2}}} \cdot e^{\left[-\frac{(-0.5-2)^2}{2\times\frac{15}{2}}\right]} = 0.096 \end{cases}$$

$$P(w_j) = \frac{|D_j|}{\sum_{i=1}^{c}|D_i|} \star \text{ (PPT 3-3)}$$

$$\because \begin{cases} p(w_1|x) = \frac{p(x|w_1)\cdot P(w_1)}{p(x)} \\ p(w_2|x) = \frac{p(x|w_2)\cdot p(w_2)}{p(x)} \end{cases} \star \text{ (PPT 3-1)}$$

$$\therefore \frac{p(w_1|x)}{p(w_2|x)} = \frac{p(x|w_1)P(w_1)}{p(x|w_2)P(w_2)} \qquad \because \begin{cases} P(w_1) = \frac{6}{10} = 0.6 \\ P(w_2) = \frac{4}{10} = 0.4 \end{cases}$$

$$= \frac{\frac{1}{\sqrt{7}}}{\frac{1}{\sqrt{\frac{15}{2}}}} \cdot \frac{0.6}{0.4} \cdot e^{\left[\frac{2.5^2}{15} - \frac{1.5^2}{14}\right]} \approx 2.0 > 1$$

$\therefore p(w_1|x) > p(w_2|x_2)$ $\therefore$ we choose w_1 the D_1 category

Final exam

1. c).

PPT2-25~29

(注意下标对应)

$\lambda_{ij} = \lambda(\partial_i | w_j)$

for ∂_1. $R(\partial_1|X) = \lambda_{11} P(w_1|X) + \lambda_{12} P(w_2|X) = \frac{0.769}{p(x)}$
$= 5 \times \frac{0.077}{p(x)} + 10 \times \frac{0.0384}{p(x)}$

for ∂_2. $R(\partial_2|X) = \lambda_{21} P(w_1|X) + \lambda_{22} P(w_2|X) = \frac{1.922}{p(x)}$
$= 10 \times \frac{0.077}{p(x)} + 30 \times \frac{0.0384}{p(x)}$

for ∂_3. $R(\partial_3|X) = \lambda_{31} P(w_1|X) + \lambda_{32} P(w_2|X) = \frac{1.6552}{p(x)}$
$= 20 \times \frac{0.077}{p(x)} + 3 \times \frac{0.0384}{p(x)}$

$X = 0.5$

注意上下题 X是否相同 不同X要重新算 $P(w_i|X)$

$\therefore R(\partial_1|X) < R(\partial_3|X) < R(\partial_2|X)$

$\therefore$ We apply ∂_1 for $X = 0.5$

d). $\because$ $D_1: \sigma_1 = 1, \mu_1 \sim N(-2, 1)$
$D_2: \sigma_2 = 2, \mu_2 \sim N(2, 1)$

PPT3-20 $\Rightarrow$ $p(\mu_1) \sim N(-2, 1)$, $p(\mu_2) \sim N(2, 1)$; $p(x|\mu_1) \sim N(\mu_1, 1)$, $p(x|\mu_2) \sim N(\mu_2, 4)$

then according to $p(\mu|D) \sim N(\mu_n, \sigma_n^2)$ PPT3-22

$$\begin{cases} \sigma_n^2 = \frac{\sigma^2 \sigma_0^2}{n\sigma_0^2 + \sigma^2} \\ \mu_n = \frac{\sigma_n^2}{\sigma^2} \sum_{k=1}^{n} X_k + \frac{\sigma_n^2}{\sigma_0^2} \mu_0 \end{cases}$$ PPT3-23

so $p(\mu_1|D_1) \sim N\left(\frac{\sigma_n^2}{1} \cdot (-12) + \frac{\sigma_n^2}{1} \cdot (-2), \frac{1 \cdot 1}{6 \cdot 1 + 1}\right)$

$p(\mu_2|D_2) \sim N\left(\frac{\sigma_n^2}{4} \cdot 8 + \frac{\sigma_n^2}{1} \cdot 2, \frac{4 \cdot 1}{4 \cdot 1 + 4}\right)$

$\Rightarrow p(\mu_1|D_1) \sim N(-2, \frac{1}{7})$ $p(\mu_2|D_2) \sim N(2, \frac{1}{2})$

then according to $p(X|D) \sim N(\mu_n, \sigma^2 + \sigma_n^2)$ PPT3-25

so $p(X|D_1) \sim N(-2, \frac{8}{7})$, $p(X|D_2) \sim N(2, \frac{9}{2})$

Under the case: $X = -0.5$ $P(w_j|x, D^*) = \frac{P(w_j) \cdot p(x|w_j, D_j)}{\sum_{i=1}^{2} P(w_i) \cdot p(x|w_i, D_i)}$ PPT3-19

$*$ 计算

$P(w_1|X=-0.5) = \frac{p(x|D_1) P(w_1)}{\sum \cdots}$

$P(w_2|X=-0.5) = \frac{p(x|D_2) P(w_2)}{\sum \cdots}$

$\Rightarrow \frac{P(w_1|X=-0.5)}{P(w_2|X=-0.5)} = \frac{p(x|D_1) P(w_1)}{p(x|D_2) P(w_2)}$

$$\Rightarrow \frac{P(w_1|X=-0.5)}{P(w_2|X=-0.5)} = \frac{\frac{3}{5} \cdot \frac{1}{\sqrt{2\pi \cdot \frac{8}{7}}} \cdot \exp\left[-\frac{1}{2} \frac{(-0.5+2)^2}{\frac{8}{7}}\right]}{\frac{2}{5} \frac{1}{\sqrt{2\pi \cdot \frac{9}{2}}} \cdot \exp\left[-\frac{1}{2} \frac{(-0.5-2)^2}{\frac{9}{2}}\right]} \approx 2 > 1$$

$\therefore w_1$

Final exam

1. e) Because use k_n-nearest neighbour to make prediction. PPT4-19

know $k_n=2$, $x=-0.5$, $p_n(x) = \frac{k_n/n}{V_n}$

$D_1 = \{-6, -4, -3, \underline{-1, 0}, 2\}$
$D_2 = \{\underline{-2, 1}, 4, 5\}$
$\Rightarrow$ $V_{n1} = 1$, $V_{n2} = 3$ $\Rightarrow$ $p_{n1}(x) = \frac{2/6}{1}$, $p_{n2}(x) = \frac{2/4}{3}$

hence,

$$\frac{P(w_1|x=-0.5)}{P(w_2|x=-0.5)} = \frac{\frac{p_{n1}(x)\cdot P(w_1)}{p(x)}}{\frac{p_{n2}(x)\cdot P(w_2)}{p(x)}} = \frac{\frac{3}{5}\times\frac{1}{3}}{\frac{2}{5}\times\frac{1}{6}} = 3 > 1$$

$\therefore$ $P(w_1|x) > P(w_2|x)$ when $x=-0.5$, we choose w_1 as decision.

---

2. a) $\because$ $\begin{pmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} \end{pmatrix}\begin{pmatrix} 3 & 3 \\ 3 & 6 \end{pmatrix} = \begin{pmatrix} 2-1 & 2-2 \\ -1+1 & -1+2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = E$

so $\Sigma_i^{-1} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} \end{pmatrix}$ when $\Sigma_1 = \Sigma_2 = \Sigma_3 = \begin{pmatrix} 3 & 3 \\ 3 & 6 \end{pmatrix}$

+ assume $\vec{x} = (x_1, x_2)$
$\vec{x}^t \Sigma_i \vec{x} = 3(x_1+x_2)^2 + 3x_2^2 > 0$
$\therefore$ positive define

b). $\because$ order principal minor determinant (顺序主子式) of $\Sigma_1\ \Sigma_2\ \Sigma_3$

is $|3| > 0$ $\begin{vmatrix} 3 & 3 \\ 3 & 6 \end{vmatrix} = 9 > 0$

so we can prove that $\Sigma_i\ (1 \le i \le 3)$ is positive definite

证明正定矩阵
① All 特征值 > 0
② 各阶顺序主子式 > 0
③ 合同矩阵 > 0

c). Case II: $\Sigma_i = \Sigma$ PPT3-50

公式正确 注意

$$\boxed{\therefore \begin{cases} g_i(\vec{x}) = \vec{w}_i^t \vec{x} + w_{i0} \\ \vec{w}_i = \Sigma^{-1}\mu_i \text{ (weight vector)} \\ w_{i0} = -\frac{1}{2}\mu_i^t \Sigma^{-1}\mu_i + \ln P(w_i) \text{ (threshold/ bias)} \end{cases}}$$

$\because$ $\mu_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\mu_2 = \begin{pmatrix} -2 \\ -1 \end{pmatrix}$, $\mu_3 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$; $P(w_1) = 0.4$, $P(w_2) = 0.4$, $P(w_3) = 0.2$

$$\begin{cases} \vec{w}_1 = \Sigma^{-1}\mu_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} & w_{10} = \ln P(w_1) = \ln\frac{2}{5} \\ \vec{w}_2 = \Sigma^{-1}\mu_2 = \begin{pmatrix} -1 \\ \frac{1}{3} \end{pmatrix} & w_{20} = -\frac{1}{2}(-2,-1)\begin{pmatrix} -1 \\ \frac{1}{3} \end{pmatrix} + \ln\frac{2}{5} = -\frac{5}{6} + \ln\frac{2}{5} \\ \vec{w}_3 = \Sigma^{-1}\mu_3 = \begin{pmatrix} 0 \\ \frac{1}{3} \end{pmatrix} & w_{30} = -\frac{1}{2}(1,2)\begin{pmatrix} 0 \\ \frac{1}{3} \end{pmatrix} = -\frac{1}{3} + \ln\frac{1}{5} \end{cases}$$

$$\begin{cases} g_1(x) = (0,0)\vec{x} + \ln\frac{2}{5} \\ g_2(x) = (-1,\frac{1}{3})\vec{x} + \ln\frac{2}{5} - \frac{5}{6} \\ g_3(x) = (0,\frac{1}{3})\vec{x} + \ln\frac{1}{5} - \frac{1}{3} \end{cases}$$

Final exam
2.d) When $x=(-2,-2)^t$

$$\Rightarrow \begin{cases} g_1(x)=(0,0)\begin{pmatrix}-2\\-2\end{pmatrix}+\ln 0.4=-0.916 \\ g_2(x)=(-1\frac{1}{3})\begin{pmatrix}-2\\-2\end{pmatrix}+\ln 0.4-\frac{5}{6}=-0.4163 \\ g_3(x)=-1+\ln 0.2=-2.6094 \end{cases} \Rightarrow g_2>g_1>g_3 \Rightarrow \omega_2 \text{ for } x(-2,-2)^T$$

different

---

3. a) $\because$ $p(x|\theta)\geq 0$ all the time

$$\int_{-\infty}^{+\infty} p(x|\theta)dx = \int_0^{+\infty} 2\theta x\cdot e^{-\theta x^2}dx \Rightarrow -\int_0^{-\infty} e^t dt = e^t\Big|_{-\infty}^{0} = 1$$

令 $t=-\theta x^2$ $\quad dt=-\theta 2x\,dx$

$0 \sim -\infty$

$\therefore p(x|\theta)$ is indeed a pdf function.

b) $p(D|\theta)=\prod_{k=1}^{n} p(x_k|\theta)$

PPT3-6 $l(\theta)=\ln p(D|\theta)=\ln\prod_{k=1}^{n} 2\theta x_k e^{-\theta x_k^2}$ (ignore $p(x|\theta)=0$ when $x<0$)

$$=\ln\left((2\theta)^n\cdot\prod_{k=1}^{n}x_k\cdot e^{-\theta\sum_{k=1}^{n}x_k^2}\right)$$

对积求导 不要写错 $\leftarrow$ $=n\cdot\ln 2\theta+\ln\prod_{k=1}^{n}x_k-\theta\cdot\sum_{k=1}^{n}x_k^2$

对θ求偏导 $\nabla_\theta l(\theta)=0 \Rightarrow \frac{2n}{2\theta}=\sum_{k=1}^{n}x_k^2 \Rightarrow \hat{\theta}=\frac{1}{\frac{1}{n}\sum_{k=1}^{n}x_k^2}$

Final exam

4 a) $\because$ know $J(w,b) = \frac{1}{2}\vec{w}^t\vec{w} - \sum_{i=1}^{4} w_i\, g(\vec{x_i})$ assume $\vec{w} = \begin{pmatrix} w_a \\ w_b \end{pmatrix}$ so $\vec{w}^t\vec{w} = w_a^2 + w_b^2$

so $\nabla J(\vec{w}, b) = \frac{1}{2}\nabla(\vec{w}^t\vec{w}) - \sum_{i=1}^{4} w_i \nabla g(\vec{x_i})$

$\nabla J(\vec{w}, b) \to \begin{bmatrix} \frac{\partial J}{\partial \vec{w}} \\ \frac{\partial J}{\partial b} \end{bmatrix}$

$$= \frac{1}{2}\begin{bmatrix} \frac{\partial J}{\partial w_a} \\ \frac{\partial J}{\partial w_b} \\ \frac{\partial J}{\partial b} \end{bmatrix} - \underbrace{\sum_{i=1}^{4} w_i \begin{bmatrix} \vec{x_i} \\ 1 \end{bmatrix}}_{} \to \text{PPT5-12}$$

$$= \frac{1}{2}\cdot 2\begin{bmatrix} w_a \\ w_b \\ 0 \end{bmatrix} - \sum_{i=1}^{4} w_i \begin{bmatrix} \vec{x_i} \\ 1 \end{bmatrix} = \begin{bmatrix} \vec{w} \\ 0 \end{bmatrix} - \sum_{i=1}^{4} w_i \begin{bmatrix} \vec{x_i} \\ 1 \end{bmatrix}$$

计算出

* 具体表述 见Quiz2

For $\vec{w_0} = (2,1)^t$, $b_0 = -1$ $\eta = 0.1$

Set $k=0$, do $k = k+1$

① $\because \nabla J(\vec{w_0}, b_0) = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} - \left(-\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} -2 \\ -7 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \\ 0 \end{bmatrix}$

$\begin{bmatrix} \vec{w} \\ 0 \end{bmatrix}$ 区分 $\begin{bmatrix} \vec{w_0} \\ b_0 \end{bmatrix}$

$k=1$ $\therefore J(\vec{w_1}, b_1) = \begin{bmatrix} \vec{w_1} \\ b_1 \end{bmatrix} = \begin{bmatrix} \vec{w_0} \\ b_0 \end{bmatrix} - \eta\, \nabla J(\vec{w_0}, b_0)$ * ← PPT3-10

$$= \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} - 0.1\begin{bmatrix} 4 \\ 8 \\ 0 \end{bmatrix} = \begin{bmatrix} 1.6 \\ 0.2 \\ -1 \end{bmatrix}$$

② $\because \nabla J(\vec{w_1}, b_1) = \begin{bmatrix} \vec{w_1} \\ 0 \end{bmatrix} - \sum_{i=1}^{n} w_i \begin{bmatrix} \vec{x_i} \\ 1 \end{bmatrix} = \begin{bmatrix} 1.6 \\ 0.2 \\ 0 \end{bmatrix} - \begin{bmatrix} -2 \\ -7 \\ 0 \end{bmatrix} = \begin{bmatrix} 3.6 \\ 7.2 \\ 0 \end{bmatrix}$

$k=2$ $\therefore J(\vec{w_2}, b_2) = \begin{bmatrix} \vec{w_2} \\ b_2 \end{bmatrix} = \begin{bmatrix} \vec{w_1} \\ b_1 \end{bmatrix} - \eta\, \nabla J(\vec{w_1}, b_1)$

$$= \begin{bmatrix} 1.6 \\ 0.2 \\ -1 \end{bmatrix} - 0.1\begin{bmatrix} 3.6 \\ 7.2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1.24 \\ -0.52 \\ -1 \end{bmatrix}$$

③ $\because \nabla J(\vec{w_2}, b_2) = \begin{bmatrix} \vec{w_2} \\ 0 \end{bmatrix} - \sum_{i=1}^{n} w_i \begin{bmatrix} \vec{x_i} \\ 1 \end{bmatrix} = \begin{bmatrix} 1.24 \\ -0.52 \\ 0 \end{bmatrix} - \begin{bmatrix} -2 \\ -7 \\ 0 \end{bmatrix} = \begin{bmatrix} 3.24 \\ 6.48 \\ 0 \end{bmatrix}$

$k=3$ $\therefore J(\vec{w_3}, b_3) = \begin{bmatrix} \vec{w_3} \\ b_3 \end{bmatrix} = \begin{bmatrix} \vec{w_2} \\ b_2 \end{bmatrix} - \eta\, \nabla J(\vec{w_2}, b_2)$

$$= \begin{bmatrix} 1.24 \\ -0.52 \\ -1 \end{bmatrix} - 0.1\begin{bmatrix} 3.24 \\ 6.48 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.916 \\ -1.168 \\ -1 \end{bmatrix}$$

$\therefore$ after three gradient descent minimize the criterion function $J(w,b)$

$\vec{w} = (0.916, -1.168)^t$ $b = -1$

$\therefore g(\vec{x}) = (0.916, -1.168)\vec{x} - 1$

b) $\because \vec{x} = (-2,-2)^t$

$\therefore g(\vec{x}) = (0.916, -1.168)\begin{pmatrix}-2\\-2\end{pmatrix} - 1 = -0.496 < 0$ → PPT5-6

$\therefore$ decide (-1) on $\vec{x} = (-2,-2)^t$

1. a) Thomas Bayes was an English statistician, philosopher and Presbyterian minister

He was born in London, England.

Prasanta Chandra Mahalanobis:

was an Indian scientist and applied statistician.

Emanuel Parzen

was an American statistician