پاسخ تمرینات یادگیری ماشین

سری اول - فاطمه حسینی ۴۰۳۴۲۲۰۶۲

۱.15) بررسی این که چگونه تعداد پارامترهای مستقل در یک چندجمله‌ای با مرتبه M و با ابعاد D به‌عنوان فضای ورودی رشد می‌کند. ابتدا مرتبه M^{th} یک چندجمله‌ای در ابعاد D را می‌نویسیم:

$$\sum_{i_1=1}^{D}\sum_{i_2=1}^{D}\cdots\sum_{i_M=1}^{D} w_{i_1 i_2 \cdots i_M} x_{i_1} x_{i_2} \cdots x_{i_M} \quad 1.133$$

$w_{i_1 i_2 \cdots i_M}$ ضرایب هستند، که تعداد ضرایب D^M اند. اما همه ضرایب به دلیل تقارن‌های تعاملی $x_{i_1} x_{i_2} \cdots x_{i_M}$ مستقل نیستند. برای محاسبه این افزونگی‌ها، چندجمله‌ای را به شکل جدیدی بازنویسی می‌کنیم:

$$\sum_{i_1=1}^{D}\sum_{i_2=1}^{i_1}\cdots\sum_{i_M=1}^{i_{M-1}} \tilde{w}_{i_1 i_2 \cdots i_M} x_{i_1} x_{i_2} \cdots x_{i_M} \quad 1.134$$

این رویکرد امکان حذف ضرایب وابسته را فراهم می‌کند و منجر به کاهش تعداد پارامترهای مستقل می‌شود. توجه داشته باشید که رابطه دقیق بین ضرایب w و $\tilde{w}$ لازم نیست صریح باشد.

تعداد پارامترهای مستقل $n(D,M)$ را می‌توان به صورت بازگشتی براساس مقادیر از مرتبه‌های پایین‌تر محاسبه کرد.

$$n(D,M)=\sum_{i=1}^{D} n(i, M-1) \quad 1.135$$

حال به کمک استقرا نشان می‌دهیم که نتیجه زیر برقرار است:

$$\sum_{i=1}^{D}\frac{(i+M-2)!}{(i-1)!\,(M-1)!}=\frac{(D+M-1)!}{(D-1)!\,M!} \quad 1.136$$

پاسخ: در 1.135 تعداد پارامترهای مستقل $n(D,M)$ که به ترتیب M ظاهر می‌شوند به‌صورت $n(D,M)=\sum_{i_1=1}^{D}\sum_{i_2=1}^{i_1}\cdots\sum_{i_M=1}^{i_{M-1}} 1$ نیز می‌توان نوشت، که دارای M تا جمله است.

$$n(D,M)=\sum_{i_1=1}^{D}\left\{\sum_{i_2=1}^{i_1}\cdots\sum_{i_M=1}^{i_{M-1}} 1\right\}$$

پس خواهیم داشت:

$$\left\{ \sum_{i_1=1}^{D} \cdots \sum_{i_M=1}^{i_{M-1}} 1 \right\} = n(i_1, M-1)$$

$$\Rightarrow n(D, M) = \sum_{i_1=1}^{D} n(i_1, M-1) \equiv (1.135)$$

که معادله است با معادله <u>1.135</u>.

برای اثبات 1.136 ابتدا (به استقراء) D را برابر مقدار یک می‌گیریم پس:

$$\sum_{i=1}^{1} \frac{(i+M-2)!}{(i-1)!(M-1)!} = \frac{(1+M-2)!}{0! \times (M-1)!} = \frac{\cancel{(M-1)!}}{1\times\cancel{(M-1)!}} = 1$$ طرف چپ معادله ⇐

طرف راست معادله ⇐

$$\frac{(D+M-1)!}{(D-1)!M!} = \frac{(1+M-1)!}{0! \times M!} = \frac{\cancel{(M)!}}{1\times\cancel{(M)!}} = 1$$

نتیجه می‌گیریم که معادله برای مقدار $D=1$ برقرار است. پس به استقراء فرض می‌کنیم معادله برای D برقرار است. حال برای مقدار D+1 اثبات می‌کنیم:

$$\sum_{i=1}^{D} \frac{(i+M-2)!}{(i-1)!(M-1)!} = \frac{(D+M-1)!}{(D-1)!M!} + \frac{(D+M-1)!}{D!(M-1)!}$$

$$= \frac{(D+M-1)!D + (D+M-1)!M}{D!M!} = \underbrace{\frac{(D+M)!}{D!M!}}$$

این قسمت دقیقاً برابر با طرف راست معادله 1.136 است وقتی که مقدار D را برابر با D+1 می‌گیریم.

پس فرض استقراء ثابت شد و حکم برقرار است. ■

<u>ادامه سوال 1.15</u>

تمرین از ما می‌خواهد تا به کمک اثبات قبل و استقراء عبارت زیر را اثبات کنیم ⇐

$$n(D, M) = \frac{(D+M-1)!}{(D-1)!M!} \qquad 1.137$$

پاسخ ⇐ به کمک استقراء معادله را در حالت $M=2$ بررسی می‌کنیم:

$$n(D,2) = \frac{(D+1)!}{(D-1)!\,2!} = \underbrace{\frac{1}{2}D(D+1)}_{\text{این معادل همان نتیجه‌ای است که در سوال 1.14 بدست آورده!}}$$

در مسأله 1.14 ما اثبات کرده‌ایم که تعداد پارامترهای ماتریس w_{ij}^{S} برابر است با $\Leftarrow \frac{D(D+1)}{2}$.

پس حالت برای $M=2$ برقرار است پس فرض می‌کنیم (به استقراء) برای $M-1$ نیز برقرار باشد. در نتیجه ⇐

$$n(D,M-1) = \frac{(D+M-2)!}{(D-1)!\,(M-1)!}$$

حالا عبارت بالا را در سمت راست معادله 1.135 جایگذاری می‌کنیم:

$$n(D,M) = \underbrace{\sum_{i=1}^{D} \frac{(D+M-2)!}{(D-1)!\,(M-1)!}}_{\text{معادل است با سمت راست معادله 1.136}}$$

و با کمک 1.136 داریم ⇐

$$\Rightarrow \quad n(D,M) = \frac{(D+M-1)!}{(D-1)!\,M!}$$

ما نشان دادیم که عبارت 1.137 برای مقادیر M برقرار است. پس براساس استقراء برای تمام مقادیر M برقرار است و حکم ثابت می‌شود. ∎

پاسخ ⇐ به کمک استقراء معادله را در حالت $M=2$ بررسی می‌کنیم:

$$n(D,2) = \frac{(D+1)!}{(D-1)!\,2!} = \underbrace{\frac{1}{2}D(D+1)}$$

این معادل همان نتیجه‌ای است که در سوال 1.14 بدست آمده!

در مسئله 1.14 ما اثبات کرده‌ایم که تعداد پارامترهای ماتریس w_{ij}^{s} برابر است با ⇐ $\frac{D(D+1)}{2}$.

پس حالت برای $M=2$ برقرار است پس فرض می‌کنیم (به استقراء) برای $M-1$ نیز برقرار باشد. در نتیجه ⇐

$$n(D,M-1) = \frac{(D+M-2)!}{(D-1)!\,(M-1)!}$$

حالا عبارت بالا را در سمت راست معادله 1.135 جایگذاری می‌کنیم:

$$n(D,M) = \underbrace{\sum_{i=1}^{D} \frac{(D+M-2)!}{(D-1)!\,(M-1)!}}_{\text{معادل است با سمت راست معادله 1.136}}$$

و با کمک 1.136 داریم ⇐

$$\Rightarrow \quad n(D,M) = \frac{(D+M-1)!}{(D-1)!\,M!}$$

ما نشان دادیم که عبارت 1.137 برای مقادیر M برقرار است. پس براساس استقراء برای تمام مقادیر M برقرار است و حکم ثابت می‌شود. ∎

۱.۱۴) این تمرین از ما می‌خواهد که یک عبارت برای تعداد کل $N(D,M)$ پارامترهای مستقل در تمام عبارات از جمله مرتبه M یک چندجمله‌ای D بعدی پیدا کنیم.

$N(D,M)$: نشان دهنده تعداد کل پارامترهای مستقل در یک چندجمله‌ای D بعدی تا مرتبه M.

یادآوری از مسأله ۱.۱۵ : $n(D,M)$ تعداد پارامترهای مستقل مرتبط با یک چندجمله‌ای از مرتبه m است و معادل است با $\Leftarrow$

$$n(D,m)=\frac{(D+m-1)!}{(D-1)!\,m!}$$

$\Downarrow$

این فرمول تعداد ترکیب‌های متمایز متغیرها را هنگام در نظر گرفتن عبارات چندجمله‌ای شمارش می‌کند و افزونگی‌های ناشی از تقارن را حذف می‌کند.

حالا تعداد کل پارامترهای مستقل تا مرتبه M از جمع روی $n(D,M)$ بدست می‌آید:

$$N(D,M)=\sum_{m=0}^{M} n(D,m)$$

پاسخ $\Leftarrow$ با جایگذاری‌ها شروع می‌کنیم:

$$N(D,M)=\sum_{m=0}^{M} n(D,m)=\sum_{m=0}^{M}\frac{(D+m-1)!}{(D-1)!\,m!}$$

$$=\sum_{i=0}^{M} n(D,m)=\underbrace{\sum_{i=0}^{M}\frac{(D+i-1)!}{(D-1)!\,i!}}$$

$\Downarrow$

این جمع مشابه فرمول ضرایب دوجمله‌ای است.

$$\sum_{i=0}^{M}\binom{D+i-1}{i}=\binom{D+M}{M}$$

از طرفی با توجه به نتایج بدست آمده از سوال 1.15 می دانیم :

$$\sum_{i=1}^{D} \frac{(i+M-2)!}{(i-1)!\,(M-1)!} = \frac{(D+M-1)!}{(D-1)!\,M!} = n(i, M-1)$$

ابتدا به استقراء برای $M=0$ بررسی می شود

$$N(D,0) = \sum_{i=0}^{0} \frac{(D+i-1)!}{(D-1)!\,i!} = \frac{(D-1)!}{(D-1)!\times 1} = 1$$

حالا به استقراء فرض می کنیم برای $M=k$ برقرار باشد. حالا برای $M=k+1$ اثبات می کنیم

$$N(D,k+1) = \sum_{i=0}^{k+1} \frac{(D+i-1)!}{(D-1)!\,i!}$$

با توجه به این که $n! \simeq n^n e^{-n}$ (تقریب استرلینگ) پس برای n های بزرگ که مقدار $N(D,M)$ رشد می کند.

$$N(D,k+1) = \frac{(D-1)!}{(D-1)!} + \frac{D!}{(D-1)!} + \frac{(D+1)!}{2(D-1)!} + \cdots + \frac{(D+k)!}{(k+1)!\,(D-1)!}$$

$$= 1 + D + 2^{-1}(D+1)D + \cdots + \frac{1}{(k+1)!}\times(D+k)\times(D+k-1)\times\cdots$$

$$\times (D) = \sum_{i=1}^{D} \frac{(i+M-2)!}{(i-1)!\,(M-1)!}$$

۱.۱۸) در این تمرین باید فرمولی برای محاسبه مساحت سطح و حجم یک کره واحد در ابعاد D با استفاده از انتگرال گاوسی و تابع گاما بدست آوریم.

S_D : مساحت سطح کره

S_v : حجم کره

مسئله به ما پیشنهاد داده که از نتیجه زیر کمک بگیریم:

$$\prod_{i=1}^{D}\int_{-\infty}^{+\infty} e^{-x_i^2}\,dx_i = S_D\int_0^{\infty} e^{-r^2} r^{D-1}\,dr \qquad \boxed{1.142}$$

باکمک تعریف $\Gamma(x) = \int_0^{\infty} u^{x-1} e^{-u}\,du \quad 1.141$ و ۱.۱۲۶ عبارات زیر را نتیجه بگیرید:

$$I = (2\pi\sigma^2)^{\frac{1}{2}} \qquad \boxed{1.126} \qquad \Rightarrow \qquad S_D = \frac{2\pi^{D/2}}{\Gamma(D/2)} \qquad \boxed{1.143}$$

و در نهایت نشان میدهیم $\Rightarrow$ $\quad v_D = \frac{S_D}{D} \qquad \boxed{1.144}$.

در پایان به کمک $\Gamma(1)=1$ ، $\Gamma(\frac{3}{2}) = \frac{\sqrt{\pi}}{2}$ ، فرمول را برای $D=2$ و $D=3$ بازنویسی میکنیم.

پاسخ $\Leftarrow$ از معادله ۱.۱۴۲ شروع میکنیم. سمت چپ معادله نشان دهنده انتگرال یک تابع گاوسی D-بعدی است در مختصات دکارتی درحالیکه سمت راست، آن را به سطح یک کره واحد در مختصات قطبی متصل میکند.

حال با کمک نتیجه ۱.۱۲۶ ، سمت چپ معادله ۱.۱۴۲ را بازنویسی میکنیم:

$$\int_{-\infty}^{+\infty} e^{-x_i^2}\,dx = \sqrt{\pi} \xrightarrow{\text{برای D بعد}} \prod_{i=1}^{D}\int_{-\infty}^{+\infty} e^{-x_i^2}\,dx_i = (\sqrt{\pi})^D = \pi^{\frac{D}{2}}$$

$(*)$

درسمت راست معادله ۱.۱۴۲ متغیر ها را تغییر می دهیم:

$$u = r^2 \Rightarrow du = 2r\,dr \Rightarrow dr = \frac{du}{2\sqrt{u}}$$

پس در انتگرال ۱.۱۴۲ با تغییر متغیر معرفی شده داریم:

$$\int_0^\infty e^{-r^2} r^{D-1}\,dr = \frac{1}{2}\int_0^\infty e^{-u} u^{(\frac{D}{2})-1}\,du = \frac{1}{2}\Gamma\left(\frac{D}{2}\right) \quad (**)$$

می دانیم

$$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u}\,du$$

$\overset{S_0}{\Rightarrow}$ طرفین معادله را (*) (**) جایگذاری می کنیم

$$\pi^{\frac{D}{2}} = S_D \frac{1}{2}\Gamma\left(\frac{D}{2}\right) \Rightarrow S_D = \frac{2\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}\right)}$$

حالا برای قسمت دوم سوال محاسبه V_D نیز داریم:

حجم V_D یک کره را می توان با ادغام مساحت سطح از شعاع ۰ تا یک پیدا کرد:

$$V_D = \int_0^1 S_D(r)\,dr \quad \text{در حالیکه} \quad S_D(r) = S_D \cdot r^{D-1}$$

$$\Rightarrow V_D = \int_0^1 \left(\frac{2\pi^{D/2}}{\Gamma(\frac{D}{2})}\right) r^{D-1}\,dr = \frac{2\pi^{D/2}}{\Gamma(\frac{D}{2})}\cdot\frac{1}{D} = \frac{2\pi^{D/2}}{D\cdot\Gamma(\frac{D}{2})} \rightarrow S_D$$

$$\Rightarrow V_D = \frac{1}{D} S_D$$

در نهایت فرمول ها را برای $D = 2, 3$ بدست می آوریم :

$$D = 2 \begin{cases} \Rightarrow S_2 = \dfrac{2\pi^{\frac{2}{2}}}{\Gamma(\frac{2}{2})} = \dfrac{2\pi}{1} = 2\pi \\[2ex] \Rightarrow V_2 = \dfrac{S_2}{2} = \dfrac{2\pi}{2} = \pi \end{cases}$$

$$D = 3 \begin{cases} \Rightarrow S_3 = \dfrac{2\pi^{\frac{3}{2}}}{\Gamma(\frac{3}{2})} = \dfrac{2\pi^{\frac{3}{2}}}{\frac{\sqrt{\pi}}{2}} = \dfrac{4\pi}{1} = 4\pi \\[2ex] \Rightarrow V_3 = \dfrac{S_3}{3} = \dfrac{4\pi}{3} \end{cases}$$

1.19) کره‌ای با شعاع a را در ابعاد D همراه با مرکز در نقطه بکشید، ابرمکعب با ضلع $2a$ را به طوری که کره در مرکز ابرمکعب را لمس کند. با استفاده از نتایج تمرین 1.18 نشان دهید که نسبت حجم کره به حجم مکعب برابر است با:

$$\frac{\text{volume of sphere}}{\text{volume of cube}} = \frac{\pi^{D/2}}{D\,2^{D-1}\,\Gamma(D/2)} \qquad 1.145$$

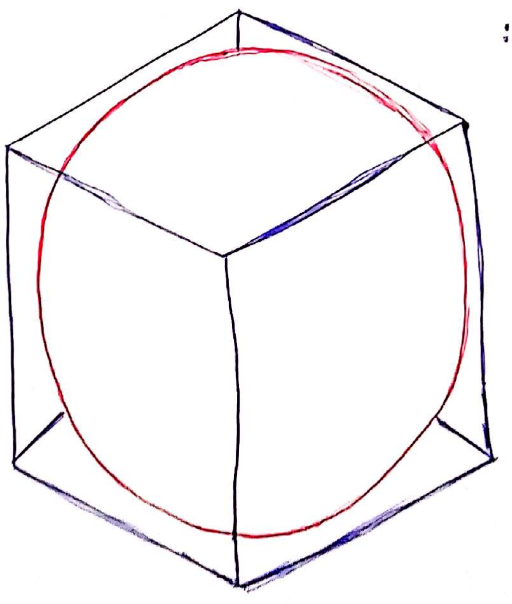
حال از فرمول استرلینگ استفاده کنید

$$\Gamma(x+1) \simeq (2\pi)^{1/2} e^{-x} x^{x+\frac{1}{2}} \qquad 1.146$$

که برای $x>1$ برقرار است، نشان دهید هرگاه $D\to\infty$، نسبت 1.145 به صفر میل می‌کند.

همچنین نشان دهید که نسبت فاصله از مرکز ابرمکعب به یکی از گوشه‌ها، تقسیم بر فاصله عمود بر یکی از اضلاع، $\sqrt{D}$ است که در نتیجه به ∞ میل می‌کند هرگاه $D\to\infty$.

از این نتایج می‌بینیم که در فضایی با ابعاد بالا، بیشتر حجم یک مکعب در تعداد زیادی گوشه‌ها متمرکز می‌شود که خود به spikes بسیار طولانی تبدیل می‌شود.



پاسخ ⇐ از تمرین 1.18 داریم:

$$v_D(r=1) = \frac{\pi^{D/2}}{\Gamma(\frac{D}{2}+1)}$$

$$v_S(a) = a^D \times \frac{\pi^{D/2}}{\Gamma(\frac{D}{2}+1)}$$

از طرفی برای حجم مکعب نیز داریم :

با اضلاع $2a$ $\Rightarrow$ $v_{cube}(2a) = (2a)^D = 2^D \cdot a^D$

محاسبه نسبت ⇐

$$\frac{v_{sphere}(a)}{v_{cube}(2a)} = \frac{a^D \cdot \frac{\pi^{\frac{D}{2}}}{\Gamma(\frac{D}{2}+1)}}{2^D \times a^D} = \frac{\pi^{\frac{D}{2}}}{2^D \times \Gamma(\frac{D}{2}+1)} \quad (*)$$

از $\Gamma(x+1) = \Gamma(x)\, x$ داریم ⇐

$$\Gamma\left(\frac{D}{2}+1\right) = \frac{D}{2}\,\Gamma\left(\frac{D}{2}\right) \quad **$$

با جایگذاری $**$ در $(*)$ داریم ⇐

$$\frac{v_{sphere}(a)}{v_{cube}(2a)} = \frac{\pi^{\frac{D}{2}}}{2^D \times D\,\Gamma(\frac{D}{2})}$$

اثبات شد.

برای قسمت بعدی هنگامی که $D \to \infty$.

نکته : برای x های بزرگ می دانیم

$$\Gamma(x) \simeq \sqrt{2\pi}\, x^{x-\frac{1}{2}}\, e^{-x}$$

$$x \to \infty \Rightarrow \Gamma(x) \to 0$$

اگر D زوج باشد ⇐ $D = 2k$ که k مقداری بزرگ است پس

$$\Gamma\left(\frac{D}{2}\right) = \Gamma(k) \qquad x > 1$$

$$\Gamma(k) \simeq \sqrt{2\pi}\; e^{-(k-1)}\, (k-1)^{k-\frac{1}{2}}$$

و اگر D فرد باشد ⇐ $D = 2k+1$

$$\Gamma\left(\frac{D}{2}\right) = \Gamma\left(k+\frac{1}{2}\right) = \Gamma\left(k-\frac{1}{2}\right) \Rightarrow \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

1.22) با توجه به یک ماتریس زیان با عناصر L_{kj}، ریسک مورد انتظار به حداقل می‌رسد، اگر برای هر x، کلاسی را انتخاب کنیم که به حداقل می‌رساند. بررسی کنید که وقتی ماتریس زیان با $L_{kj}=1-I_{kj}$ که I_{kj} اعضای ماتریس همانی اند. این به معیار انتخاب کلاسی که بیشترین احتمال پسین را دارد کاهش می‌دهد. تفسیر این شکل از ماتریس زیان چیست؟

پاسخ ⇐ در این تمرین هدف ما این است که نشان دهیم که به حداقل رساندن ریسک مورد انتظار با یک ماتریس زیان خاص منجر به انتخاب کلاس با بالاترین احتمال پسین می‌شود.

ماتریس زیان در این سوال $L_{kj}=1-I_{kj}$

$$\begin{cases} I_{kj}=1 & \text{if} \quad k=j \\ I_{kj}=0 & \text{if} \quad k\neq j \end{cases}$$

می‌دانیم ریسک مورد انتظار برای کلاس j و ورودی X به صورت زیر تعریف می‌شود:

$$Risk(j|X)=\sum_k L_{kj}\cdot P(C_k|X)$$

حالا در این معادله به جای L_{kj} از $1-I_{kj}$ استفاده می‌کنیم پس:

$$Risk(j|X)=\sum_k (1-I_{kj})\,P(C_k|X)$$

sum بالا را باز می‌کنیم و به صورت زیر بازنویسی می‌کنیم:

$$Risk(j|X)=(1-I_{jj})\,P(C_j|X)+\sum_{k\neq j}(1-I_{jk})\,P(C_k|X)$$

از اون جایی که $I_{jj}=1$ به دلیل تعریف ماتریس همانی و $I_{jk}=0$ وقتی $j\neq k$ داریم:

$$Risk(j|X)=0\times P(C_j|X)+\sum_{k\neq j}P(C_k|X)$$

$$\Rightarrow \quad Risk(j|X)=\sum_{k\neq j}P(C_k|X) \quad (*)$$

حالا با استفاده از احتمال پسین و دانستن این نکته که مجموع احتمالات برابر یک می‌شود داریم:

$$\sum_k P(C_k|X) = 1$$

سپس (*) را دوباره با توجه به نکته بالا بازنویسی می‌کنیم:

$$Risk(j|X) = 1 - P(C_j|X)$$

واضح است که برای min کردن ماتریس زیان باید کلاسی انتخاب شود که $P(C_j|X)$ به صورت max باشد.

۱.۲۳) زمانی که ماتریس زیان عمومی و احتمالات قبلی کلی برای کلاس ها وجود دارد، معیار به حداقل رساندن زیان مورد انتظار را بدست آورید.

پاسخ: زیان مورد انتظار $E[L]$ به صورت زیر تعریف می‌شود:

$$E[L] = \sum_k \sum_j \int_{R_j} L_{kj} \, P(X, C_k) \, dx$$

L_{kj} ⊛ زیان است وقتی یک ورودی X به عنوان کلاس j طبقه‌بندی می‌شود در حالی که کلاس درست C_k است.

⊛ R_j نیز ناحیه مورد نظر برای کلاس C_j است.

هدف ما انتخاب R_j یا همان مناطق تصمیم‌گیری است برای به حداقل رساندن زیان مورد انتظار در تمام ورودی ها.

برای هر بردار ورودی X ما تصمیم می‌گیریم کدام کلاس j را به طور مستقل به آن اختصاص دهیم. برای به حداقل رساندن زیان مورد انتظار کلی، باید به به حداقل رساندن سهم برای هر X خاص تمرکز شود:

$$\sum_k L_{kj} \, P(X, C_k)$$

از طرفی $P(X, C_k) = P(C_k|X)P(X)$ پس داریم

$$\sum_k L_{kj} \, P(C_k|X) P(X)$$

از آن جایی که $P(X)$ ثابت و مثبت است برای هر X داده شده، پس قابل چشم پوشی است و به حداقل رساندن عبارت بالا معادل است با به حداقل رساندن

$$\sum_k L_{kj} \, P(C_k|X)$$

با اعمال قاعده بیز داریم: $P(C_k|X) = \dfrac{P(X|C_k)P(C_k)}{P(X)}$ که $P(X|C_k)$ چگالی کلاس است و $P(C_k)$ احتمال پیشین برای کلاس C_k است. پس

$$P(X) = \sum_k P(X|C_k) P(C_k)$$

از آن جایی که $P(X)$ برای همه کلاس ها یکسان است ما می‌توانیم روی به حداقل رساندن $\sum_k L_{kj} \, P(X|C_k) P(C_k)$ تمرکز کنیم. برای min کردن زیان مورد انتظار هر ورودی X را به کلاس j مورد نظر آن اختصاص می‌دهیم:

$$j = \underset{j}{\operatorname{argmin}} \sum_k L_{kj} \, P(C_k|X)$$

۱.۲۴) یک مسئله طبقه بندی را در نظر بگیرید که در آن زیان ایجاد شده زمانی که بردار ورودی از کلاس C_k به عنوان کلاس C_j طبقه بندی می‌شود، توسط ماتریس زیان L_{kj}ها داده می‌شود، که ضرر وارد شده در انتخاب گزینه رد λ است. معیار تصمیم گیری را پیدا کنید که حداقل ضرر مورد انتظار را ایجاد کند. بررسی کنید که وقتی ماتریس ضرر با $L_{kj}=1-I_{kj}$ ارائه می‌شود، رابطه بین λ و آستانه رد θ چیست؟

پاسخ: بردار x به کلاس C_k با احتمال $P(C_k|X)$ متعلق است. اگر به اشتباه به کلاس C_j تعلق گیرد ما شاهد یک زیان مورد انتظار هستیم که با $\sum_k L_{kj} \cdot P(C_k|X)$ نمایش داده می‌شود در حالی که اگر گزینه رد را انتخاب کنیم متحمل ضرر خواهیم شد. بنابراین اگر

$$j = \underset{J}{\arg\min} \sum_k L_{kJ} P(C_k|X)$$

سپس min سازی می‌کنیم این زیان را:

$$\text{class } j \quad \text{if } \min_J \sum_k L_{kJ} P(C_k|X) < \lambda\,;$$ (حالت اول

در غیر این صورت رد میشه (حالت دوم

برای یک ماتریس زیان $L_{kj}=1-I_{kj}$ داریم: $\sum_k L_{kJ} P(C_k|X) = 1 - P(C_J|X)$

و در نتیجه رد می‌کنیم مگر این که کوچکترین مقدار $1-P(C_j|X)$ از λ کمتر باشد یا برابر است اگر بزرگترین مقدار $P(C_j|X)$ از $1-\lambda$ کمتر باشد. در مقدار استاندارد آستانه رد ما رد می‌کنیم اگر بزرگترین احتمال پیشین از θ کمتر باشد. در نتیجه این دو معیار برای رد برابرند با $\theta = 1-\lambda$.