به نام خدا

دانشکده‌ی علوم و مهندسی کامپیوتر

یادگیری ماشین

پاسخ‌های تکلیف اول – پاییز ۱۳۹٤

این پاسخ‌ها نمونه‌هایی هستند که حاوی نکات اصلی می‌باشند و لزوما تنها پاسخ‌های درست نیستند.

۱- الف- تابع $p(x|\theta)$ به صورت کاملا صریح در جدول مشخص شده است:

$$p(x=0|\theta)=\begin{cases}0.9 & \theta=1\\ 0.6 & \theta=2\\ 0.2 & \theta=3\end{cases}$$

و طبیعتا:

$$p(x=1|\theta)=\begin{cases}0.1 & \theta=1\\ 0.4 & \theta=2\\ 0.8 & \theta=3\end{cases}$$

این تابع نشان‌دهنده‌ی این است که اگر شما بدانید با چه حریفی روبررو خواهید شد، آنگاه احتمال برد یا باخت شما چقدر است. در واقع دانستن حریف، توزیع برد و باخت شما را مشخص می‌کند.

برای محاسبه‌ی $p(x)$ نیازمند تابع $p(\theta)$ هستیم که به صورت دقیق در جدول ذکر شده است:

$$p(\theta)=\begin{cases}0.3 & \theta=1\\ 0.5 & \theta=2\\ 0.2 & \theta=3\end{cases}$$

این تابع نشان‌دهنده‌ی این است که چقدر امکان دارد که با هر یک حریف‌ها مواجه شوید.

حال برای محاسبه‌ی $p(x)$ می‌توانیم از رابطه‌ی زیر استفاده نماییم:

$$p(x)=\sum_{i=1}^{3}p(x|\theta=i)\,p(\theta=i)$$

در نتیجه خواهیم داشت:

$$p(x)=\begin{cases}0.61 & x=0\\ 0.39 & x=1\end{cases}$$

مقادیر این تابع بازگوکننده‌ی احتمال برد یا باخت شما به طور کلی (مستقل از اینکه حریف شما کیست) است.

حال با داشتن این توابع می‌توانیم $p(\theta|x)$ را با به‌کارگیری قانون بیز محاسبه نماییم:

$$p(\theta=1|x)=\begin{cases}0.44 & x=0\\ 0.08 & x=1\end{cases}$$

$$p(\theta=2|x)=\begin{cases}0.49 & x=0\\ 0.51 & x=1\end{cases}$$

$$p(\theta=3|x)=\begin{cases}0.07 & x=0\\ 0.41 & x=1\end{cases}$$

حال اگر شما نتیجه مسابقه را بدانید، تابع $p(\theta|x)$ مشخص می‌کند که با چه احتمالی با هر یک از رقبا روبه‌رو بوده‌اید. این تابع به نوعی استنتاج می‌کند. در واقع با دانستن نتیجه‌ی مسابقه، توزیع احتمال روبه‌رو بودن با حریف‌ها را مشخص می‌کند.

ب- با استفاده از تابع $p(\theta|x)$ می‌توان متوجه شد که در صورتی که x=1 باشد (یعنی مسابقه را برده باشید) احتمالا با قهرمان دوم مسابقه داده‌اید. زیرا احتمال $p(\theta=2|x=1)$ بیشتر از $p(\theta=1|x=1)$ و $p(\theta=3|x=1)$ است.

ج- واضح است که حتی اگر جایزه و ضرر هر دو حتی یکسان هم باشند، بازهم شرکت در این مسابقه به‌صرفه نیست، زیرا

$$p(x=0)>p(x=1)$$

در اینجا حتی ضرر بیشتر از جایزه است و قطعا نباید در مسابقه شرکت کرد! از نظر ریاضی باید امید متغیر تصادفی زیر را حساب نماییم:

$$g(x)=\begin{cases}-2000000 & x=0\\ 1000000 & x=1\end{cases}$$

این متغیر تابعی از x (برد یا باخت) است و در نتیجه امید آن به صورت زیر خواهد بود:

$$E[g(x)]=\sum_{i=0}^{1}g(x=i)\,p(x=i)=-1220000+390000=-830000$$

که به این معنی است که شما اگر این مسابقه را بارها تکرار نمایید، به طور متوسط در هر بار ۸۳۰ هزار تومان ضرر خواهید کرد! پس با این اطلاعات، هوشمندانه نیست که در این مسابقه شرکت کنید.

۲- الف- با توجه به توضیحات مسئله، متغیرها i.i.d هستند و در نتیجه تابع likelihood به صورت زیر خواهد بود:

$$l(\lambda|k)=\prod_{i=1}^{n}p(k_i|\lambda)=\prod_{i=1}^{n}\frac{\lambda^{k_i}e^{-\lambda}}{k_i!}$$

با گرفتن لگاریتم در مبنای e از طرفین خواهیم داشت:

$$L(\lambda|k)=\sum_{i=1}^{n}\ln\left(\frac{\lambda^{k_i}e^{-\lambda}}{k_i!}\right)=\sum_{i=1}^{n}\left(\ln\left(\lambda^{k_i}\right)+\ln\left(e^{-\lambda}\right)-\ln\left(k_i!\right)\right)=\sum_{i=1}^{n}\left(k_i\ln\lambda-\lambda-\ln\left(k_i!\right)\right)=\ln\lambda\sum_{i=1}^{n}k_i-n\lambda-n\ln\left(k_i!\right)$$

از این تابع نسبت به λ مشتق گرفته و مساوی صفر قرار می‌دهیم:

$$\frac{\partial L(\lambda \mid k)}{\partial \lambda} = \frac{\sum_{i=1}^{n} k_i}{\lambda} - n = 0$$

در نتیجه:

$$\lambda = \frac{\sum_{i=1}^{n} k_i}{n}$$

یعنی پارامتر λ باید برابر با میانگین داده‌ها باشد.

ب- فرض می‌کنیم که $\tilde{\lambda}$ یک متغیر تصادفی است که در واقع تخمین λ از روی یک دیتاست حاوی n نمونه است که همگی از یک توزیع پواسون با پارامتر λ استخراج شده‌اند. برای اینکه بایاس وجود نداشته باشد، باید امید متغیر تصادفی $\tilde{\lambda}$ برابر با λ باشد:

$$E\left[\tilde{\lambda}\right] = E\left[\frac{\sum_{i=1}^{n} k_i}{n}\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[k_i\right]$$

می‌دانیم که تمامی متغیرهای تصادفی k_i از یک توزیع پواسون با پارامتر λ استخراج شده‌اند. به علاوه میانگین توزیع پواسون نیز همان λ است. در نتیجه:

$$E\left[\tilde{\lambda}\right] = \frac{1}{n}\sum_{i=1}^{n} \lambda = \frac{1}{n} n\lambda = \lambda$$

پس این تخمین دارای بایاس نیست.

ج- برای این اثبات به صورت زیر عمل می‌نماییم:

$$Var\left(\tilde{\lambda}\right) = Var\left(\frac{\sum_{i=1}^{n} k_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var\left(k_i\right)$$

می‌دانیم که تمامی متغیرهای تصادفی k_i از یک توزیع پواسون با پارامتر λ استخراج شده‌اند. به علاوه واریانس توزیع پواسون نیز همان λ است. در نتیجه:

$$Var\left(\tilde{\lambda}\right) = \frac{1}{n^2} n\lambda = \frac{\lambda}{n}$$

د- سازگاری یک توزیع به این معنی است که بایاس و واریانس آن با افزایش n به سمت صفر بروند. در اینجا داریم:

۱) طبق پاسخ بخش ب بایاس این تخمین صفر است.

۲) طبق پاسخ بخش ج واریانس این تخمین با افزایش n به سمت صفر می‌رود.

در نتیجه این تخمین، یک تخمین سازگار است.

ه- در این بخش دانشی در مورد پارامتر λ از پیش داریم (Prior Knowledge)، در نتیجه به جای تکنیک MLE باید از تکنیک MAP برای تخمین زدن استفاده نماییم:

$$p(\lambda|k)=\frac{p(\lambda)\,p(k|\lambda)}{p(x)}=\frac{p(\lambda)\prod_{i=1}^{n}p(k_i|\lambda)}{p(x)}$$

از طرفین لگاریتم در مبنای e می‌گیریم:

$$\ln p(\lambda|k)=\ln p(\lambda)+\sum_{i=1}^{n}\ln p(k_i|\lambda)-\ln p(x)$$

که پس از ساده‌سازی خواهیم داشت:

$$\ln p(\lambda|k)=\ln\lambda-\ln 100+\ln\lambda\sum_{i=1}^{n}k_i-n\lambda-n\ln(k_i!)-\ln p(x)$$

از این رابطه مشتق گرفته و مساوی صفر قرار می‌دهیم:

$$\frac{\partial\ln p(\lambda|k)}{\partial\lambda}=\frac{1}{\lambda}+\frac{\sum_{i=1}^{n}k_i}{\lambda}-n=0$$

در نتیجه خواهیم داشت:

$$\lambda=\frac{1+\sum_{i=1}^{n}k_i}{n}$$

از طرفی مرز بازه‌ی مقادیر λ نیز جز نقاط بحرانی هستند و ممکن است که مقدار بهینه در آنها رخ دهد. پس مقدار بهینه‌ی λ به صورت زیر است:

$$\lambda=\underset{\lambda_i\in S}{\arg\max}\left(p(\lambda_i|k)\right)$$

به طوری که:

$$S=\left\{\frac{1+\sum_{i=1}^{n}k_i}{n},5,15\right\}$$

پس تنها می‌توان گفت مقدار بهینه یکی از این سه مقدار است، ولی جواب نهایی وابسته به داده‌ها است.

۳- ظاهر این سوال ممکن است این‌طور به نظر برسد که بسیار مشابه سوال قبلی است، اما برای اثبات بایاس، کاملا تفاوت دارد.

ابتدا با استفاده از روش Maximum Likelihood مقدار بهینه‌ی پارامتر λ را تخمین می‌زنیم:

$$l(\lambda \mid x) = \prod_{i=1}^{n} p(x_i \mid \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}$$

با گرفتن لگاریتم در مبنای e از طرفین خواهیم داشت:

$$L(\lambda \mid x) = \sum_{i=1}^{n} \ln\left(\lambda e^{-\lambda x_i}\right) = \sum_{i=1}^{n}\left(\ln(\lambda) - \lambda x_i\right) = n\ln(\lambda) - \lambda\sum_{i=1}^{n} x_i$$

حال از این تابع نسبه به λ مشتق گرفته و مساوی صفر قرار می‌دهیم:

$$\frac{\partial L(\lambda \mid x)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

در نتیجه:

$$\lambda = \frac{n}{\sum_{i=1}^{n} x_i}$$

تاکنون مقدار پارامتر بهینه را با استفاده از تکنیک MLE برای توزیع نمایی به دست آوردیم. حال تازه به بخش اصلی سوال یعنی اثبات بایاس این تخمین رسیدیم. فرض می‌کنیم که $\tilde{\lambda}$ یک متغیر تصادفی است که در واقع تخمین λ از روی یک دیتاست حاوی n نمونه است که همگی از یک توزیع پواسون با پارامتر λ استخراج شده‌اند. در ادامه قصد داریم تا امید $\tilde{\lambda}$ را حساب نماییم:

$$E\left[\tilde{\lambda}\right] = E\left[\frac{n}{\sum_{i=1}^{n} x_i}\right] = nE\left[\frac{1}{\sum_{i=1}^{n} x_i}\right]$$

اولین نکته این است که نمی‌توانیم مانند روش قبلی عمل نماییم و سیگما را از E خارج نماییم، زیرا سیگما در مخرج است. اما می‌دانیم که عبارت داخل E خودش یک متغیر تصادفی است که بر حسب اعضای خارج شده از توزیع (x_i ها) هستند. با تخمین توزیع این متغیر می‌توانیم امید آن را حساب نماییم.

ابتدا توزیع مجموع x_i ها را به دست آوریم. این متغیر را $\overline{x}_n$ می‌نامیم. از طرفی می‌دانیم که مجموع تعداد محدودی از متغیرهای تصادفی از یک توزیع نمایی با پارامتر λ، از توزیع ارلنگ پیروی می‌کند که پارامترهای آن n و λ هستند:

$$\bar{x}_n = \sum_{i=1}^{n} x_i ,\ \mathrm{p}\left(x_i\right) = \lambda e^{-\lambda x_i} \Rightarrow p\left(\bar{x}_n\right) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}$$

پس مسئله به صورت زیر تغییر می‌کند:

$$E\left[\tilde{\lambda}\right] = nE\left[\frac{1}{\bar{x}_n}\right]$$

حال باید امید متغیر تصادفی $\frac{1}{\bar{x}_n}$ را حساب نماییم که در واقعی تابعی از $\bar{x}_n$ است. اینجا دو راه وجود دارد. یکی اینکه توزیع $\frac{1}{\bar{x}_n}$ را بـه دست آوریم و سپس از آن امید بگیریم که راه نسبتا سختی است. راه دوم مسئله را مستقیم حل می‌کند. در راه دوم از قضیه زیر استفاده می‌کنیم:

$$E\left[g\left(x\right)\right] = \int_{-\infty}^{\infty} g\left(x\right) f\left(x\right) dx$$

پس خواهیم داشت: (بازه‌ی انتگرال از صفر شروع می‌شود زیرا مقدار متغیر $\bar{x}_n$ امکان ندارد که کمتر از صفر باشد)

$$E\left[\tilde{\lambda}\right] = nE\left[\frac{1}{\bar{x}_n}\right] = \int_0^{\infty} \frac{1}{\bar{x}_n} p\left(\bar{x}_n\right) d\bar{x}_n = \int_0^{\infty} \frac{\lambda^n \bar{x}_n^{\,n-1} e^{-\lambda \bar{x}_n}}{\bar{x}_n (n-1)!} d\bar{x}_n = \frac{\lambda^n}{(n-1)!} \int_0^{\infty} \bar{x}_n^{\,n-2} e^{-\lambda \bar{x}_n} d\bar{x}_n$$

برای حل این انتگرال باید n-2 بار از روش جز به جز استفاده نمایید. اما نگران نباشید، این کار قبلا انجام شده است و یک فرمـول کلـی دارد:

$$\int_0^{\infty} x^n e^{-ax} dx = \frac{n!}{a^{n+1}}$$

در نتیجه:

$$E\left[\tilde{\lambda}\right] = \frac{\lambda^n}{(n-1)!} \int_0^{\infty} \bar{x}_n^{\,n-2} e^{-\lambda \bar{x}_n} d\bar{x}_n = \frac{\lambda^n}{(n-1)!} \times \frac{(n-2)!}{\lambda^{n-1}} = \frac{\lambda}{n-1}$$

پس این تخمین دارای بایاس است، زیرا امید تخمین دقیقا برابر با λ نشد.

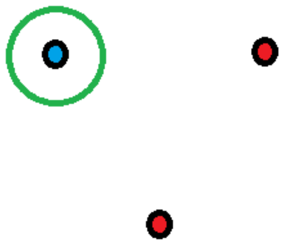
❖ یک نکته جالب این است که اگر از فرم دیگر توزیع نمایی استفاده نماییم (مانند زیر)، تخمین پارامتر β آن دارای بایاس نخواهد بود!

$$f\left(x\,|\,\beta\right)=\begin{cases}\dfrac{1}{\beta}e^{-\frac{x}{\beta}} & x\geq 0\\ 0 & x<0\end{cases}$$

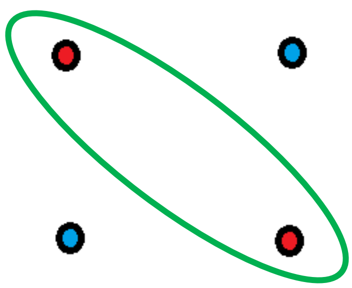
۴- از یک داده شروع می‌کنیم و به ترتیب تعدادشان را زیاد می‌نماییم تا جایی که امکان Shatter شدن وجود نداشته باشد. برای سادگی کلاس‌ها را با نام‌های آبی و قرمز ذکر می‌نماییم.

یک حقه مناسب این است که هر بار داده‌هایتان را روی یک چندضلعی منتظم فرض نمایید. زیرا با این کار بسیاری از حالت‌های برچسب‌زنی مانند هم هستند و در نتیجه نیاز به بررسی همه حالات نیست.

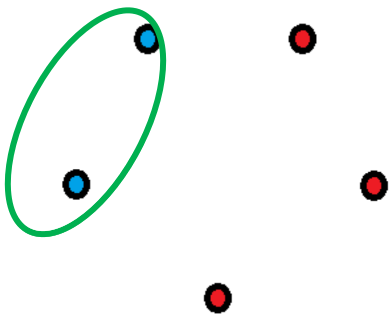
- **۱ داده:** در این حالت بیضی تنها کافی است که یک نقطه را دربربگیرد و قطعا توانایی این کار را دارد.
- **۲ داده:** بیضی می‌تواند هر داده (اگر یکی قرمز و دیگری آبی باشد) یا هر دو داده (اگر هر دو در یک کلاس باشند) را در بر بگیرد.
- **۳ داده:** این حالت دارای ۸ حالت برچسب زنی است. سخت‌ترین حالت وجود ۱ آبی و ۲ قرمز (یا برعکس) است که می‌توان بیضی را دور آبی در نظر گرفت و در نتیجه داده‌ها را جدا کرد.

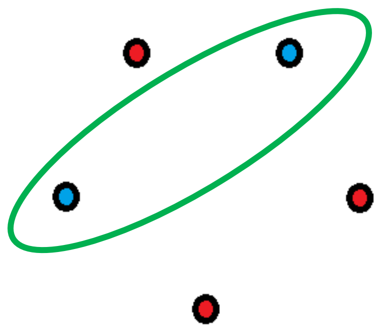


- **۴ داده:** داده‌ها را روی چهارگوش مربع در نظر می‌گیریم. اگر یک آبی و سه قرمز (یا برعکس) داشته باشیم، به سادگی بیضی را دور آبی در نظر می‌گیریم. اما بدترین حالت وقتی است که دو سر قطرها با هم در یک کلاس باشند. این حالت نیز توسط بیضی دوران-یافته قابل جداسازی است.

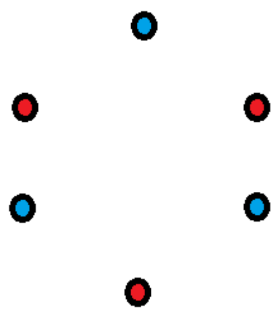


- **۵ داده:** در این حالت نقاط روی پنج‌ضلعی منتظم در نظر گرفته می‌شوند. اگر یک آبی و چهار قرمز (یا برعکس) وجود داشته باشند که باز به سادگی بیضی دور یک آبی قرار می‌گیرد و جداسازی انجام می‌شود. اما حالت مشکل وقتی است که دو آبی و سه قرمز (یا برعکس) داشته باشیم. در این حالت، یا دو آبی کنار هم هستند یا جدا از هم، که هر دوی آنها مانند شکل زیر توسط بیضی جداپذیر هستند:





❖ **۶ داده:** در این حالت، متوجه می‌شویم که با در نظر گرفتن داده‌ها روی راس‌های شش ضلعی منتظم، در همه‌ی حالات بیضی توانایی جداسازی آنها را ندارد. به عنوان مثال در حالت زیر بیضی قادر به جداسازی آبی و قرمز از یکدیگر نیست:



در نتیجه بعد VC یک بیضی، حداقل ۵ است.

۵- با توجه به اینکه هزینه ی انتخاب اشتباه برای هر کلاس می‌تواند متفاوت باشد، تابع جداساز را بر مبنای کمینه کردن ریسک انتخاب تعیین می کنیم:

$$g_i(x) = -R(\alpha_i \mid x)$$

اگر قرار باشد مرز جداساز در نقطه ی $x = L$ قرار گیرد لازم است:

$$g_1(x = L) = g_2(x = L) \quad => \quad R(\alpha_1 \mid x = L) = R(\alpha_2 \mid x = L)$$

که،

$$R(\alpha_i \mid x) = \sum_{k=1}^{K} \lambda_{ik} P(C_k \mid x)$$

در اینجا $k = 2$ و $\lambda_{12} = 0.5$ و $\lambda_{21} = m$ است. بدین ترتیب با جایگذاری و کمی ساده‌سازی مقدار m بدین شکل بدست می آید:

$$m = \frac{1}{3}\exp(2L - 6)$$

۶- الف) معادله‌ی صفحه‌ی جداساز برای دو کلاس با ماتریس کواریانس مشترک در بخش (ج) محاسبه شده است.

ب)

ج) برای دسته‌بندی یک نمونه‌ی جدید در یک دسته‌بندی بیزی دو کلاسه، مقدار تابع جداساز برای هر کلاس در آن نقطه محاسبه شده و نمونه به کلاسی که مقدار بیشتری دارد تعلق می گیرد. اما برای نقاط روی مرز جداساز دو کلاس، احتمال تعلق به هر دو کلاس یکسان است. این مرز جداساز برای داده‌های ۳ بعدی یک صفحه است. در این بخش هدف پیدا کردن معادله صفحه مرز جداساز است تا بتوانیم آن را رسم نماییم.

معادله‌ی صفحه به صورت زیر تعریف می شود:

$$ax_1 + bx_2 + cx_3 = d$$

بنابراین مجهولات مسأله a و b و c و d هستند.

می‌دانیم که احتمال تعلق هر نقطه روی مرز جداساز به هر کلاس برابر ½ و یا برابر احتمال تعلق کلاس دیگراست. پس:

$$P(C_i \mid x) = 1/2 \ \text{ OR } \ P(C_1 \mid x) = P(C_2 \mid x)$$

$$\frac{P(x|C_i)P(C_i)}{P(x)} = 1/2 \ \text{ OR } \ \frac{P(x|C_1)P(C_1)}{P(x)} = \frac{P(x|C_2)P(C_2)}{P(x)} \Rightarrow P(x \mid C_1)P(C_1) = P(x \mid C_2)P(C_2)$$

با گرفتن لگاریتم از طرفین و در نظر گرفتن ماتریس کواریانس مشترک، داریم:

$$-1/2\,(x-m_1)^T S^{-1}(x-m_1) + \log\hat{p}(C_1) = -1/2\,(x-m_2)^T S^{-1}(x-m_2) + \log\hat{p}(C_2) \quad (۶.۱)$$

که،

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} m_1 = \begin{bmatrix} m_{11} \\ m_{21} \\ m_{31} \end{bmatrix} m_2 = \begin{bmatrix} m_{12} \\ m_{22} \\ m_{32} \end{bmatrix} S^{-1} = \begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{bmatrix} m_d = (m_2 - m_1) = \begin{bmatrix} m_{d1} = (m_{12} - m_{11}) \\ m_{d2} = (m_{22} - m_{21}) \\ m_{d3} = (m_{32} - m_{31}) \end{bmatrix}$$

و S تخمین ماتریس کواریانس مشترک، m تخمین بردار میانگین و p(C) تخمین دانش پیشین هستند. S از رابطه ی زیر محاسبه می شود:

$$S = \sum_i \hat{p}(C_i) S_i$$

در ادامه با ساده‌سازی رابطه‌ی ۶٫۱ معادله ی صفحه مورد نظر بدست می آید.

$$=>(x-m_1)^T S^{-1}(x-m_1)-(x-m_2)^T S^{-1}(x-m_2)=2(\log\hat{p}(C_1)-\log\hat{p}(C_2))$$

$$=>2[x^T S^{-1}m_2-x^T S^{-1}m_1]=2(\log\hat{p}(C_1)-\log\hat{p}(C_2))+m_2^T S^{-1}m_2-m_1^T S^{-1}m_1$$

$$=>x^T S^{-1}(m_2-m_1)=1/2\,[2(\log\hat{p}(C_1)-\log\hat{p}(C_2))+m_2^T S^{-1}m_2-m_1^T S^{-1}m_1]$$

$$=>[S_{11}m_{d1}+S_{12}m_{d2}+S_{13}m_{d3}]x_1+[S_{21}m_{d1}+S_{22}m_{d2}+S_{23}m_{d3}]x_2+[S_{31}m_{d1}+S_{32}m_{d2}+S_{33}m_{d3}]x_3= 1/2\,[2(\log\hat{p}(C_1)-\log\hat{p}(C_2))+m_2^T S^{-1}m_2-m_1^T S^{-1}m_1] \quad (۶.۲)$$

که،

$$a=S_{11}m_{d1}+S_{12}m_{d2}+S_{13}m_{d3}$$

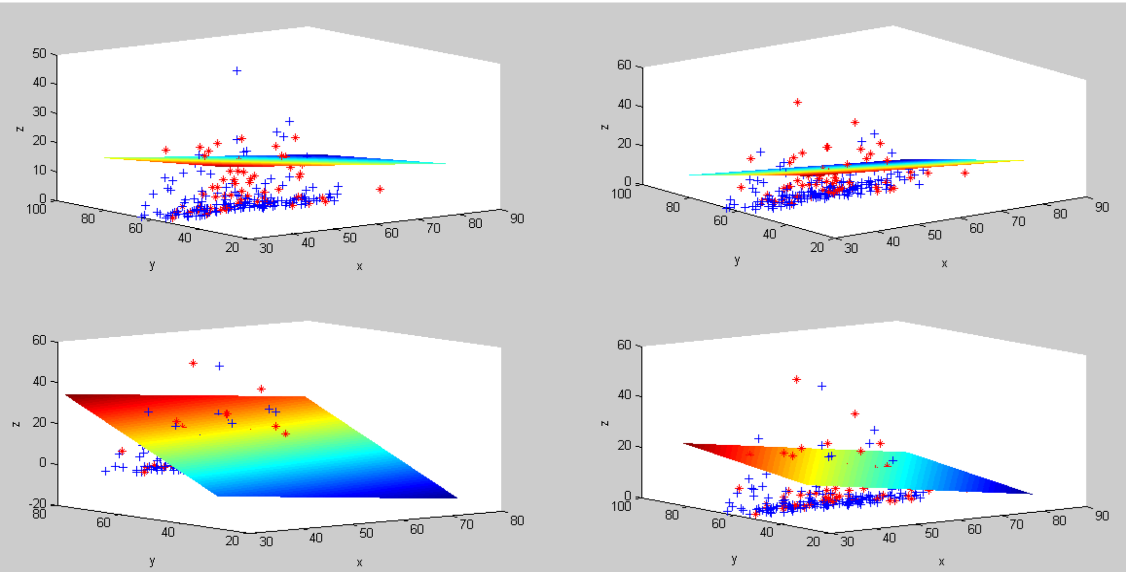
$$b=S_{21}m_{d1}+S_{22}m_{d2}+S_{23}m_{d3}$$

$$c=S_{31}m_{d1}+S_{32}m_{d2}+S_{33}m_{d3}$$

$$d=1/2\,[2(\log\hat{p}(C_1)-\log\hat{p}(C_2))+m_2^T S^{-1}m_2-m_1^T S^{-1}m_1]$$

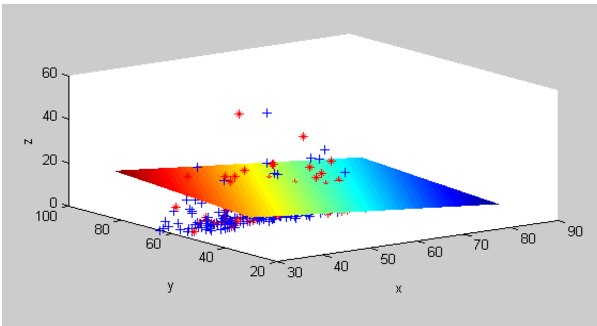
حال می‌توان معادله صفحه بدست آمده (معادله ۶,۱) را با نمونه برداری رسم نمود.

شکل ۶,۱ جداساز بدست آمده در هر بار دسته‌بندی را برای نمونه‌های آزمایش نشان می‌دهد.

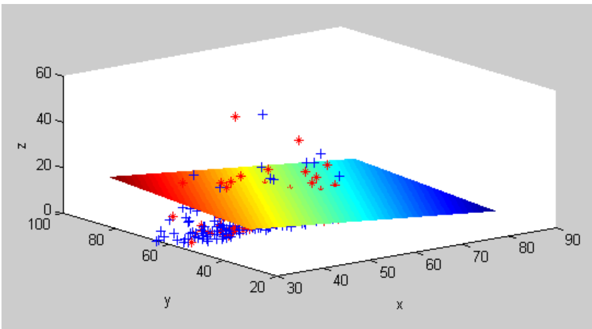


شکل ۶,۱

د) شکل ۶,۲ جداساز بدست آمده در حالت استفاده از تمام نمونه‌ها برای آموزش و شکل ۶,۳ جداساز بدست آمده از میانگین ۴ مدل محاسبه شده در ۴ بار تقسیم‌بندی داده‌ها را نشان می دهد. مقادیر پارامتر های مدل ها در جدول ۶,۱ آمده است. مشاهده می‌شود که میانگین مدل های بدست آمده شباهت زیادی به مدل بدست آمده از کل داده‌ها دارد.



شکل ۶,۲



شکل ۶,۳

داده‌ها	a	b	c	d
دسته ی اول	۰,۰۳۰۴	-۰,۰۴۹۶	۰,۱۰۶۸	۰,۰۳۴۶
دسته ی دوم	۰,۰۰۸۹	۰,۲۶۵	۰,۰۸۵۱	۳,۵۶۹۸
دسته ی سوم	۰,۰۱۵۶	۰,۰۰۳۷	۰,۱۰۸۹	۲,۶۳۹۴
دسته ی چهارم	۰,۰۲۷۱	-۰,۰۱۶۷	۰,۱۰۵۰	۲,۰۱۳۴
میانگین ۴ مدل	۰,۰۲۰۵	-۰,۰۰۹۰	۰,۱۰۱۵	۲,۰۶۴۳
همه ی نمونه ها	۰,۰۲۰۱	-۰,۰۰۸۷	۰,۱۰۰۶	۲,۰۵۰۷

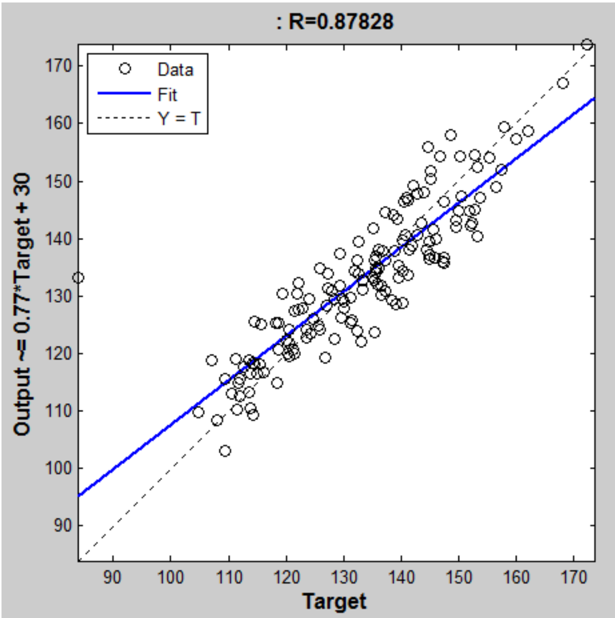
جدول ۶,۱

۷- الف) می‌توان با استفاده از همبستگی خطی بین ویژگی قیمت و سایر ویژگی‌ها این رتبه‌بندی را انجام داد. (جدول ۷,۱)

ویژگی	همبستگی
اول	۰.۸۴۱۲
سوم	۰.۷۶۲۴
چهارم	۰.۶۰۶۴

جدول ۷,۱

ب) مقدار همبستگی بین ویژگی قیمت و تخمین آن با استفاده از هر ۳ ویژگی: ۰,۸۷۸۳ و میانگین مربعات خطا: ۴۸,۶۵۷۹ است. شکل ۷,۱ مدل بدست آمده را نمایش می دهد.

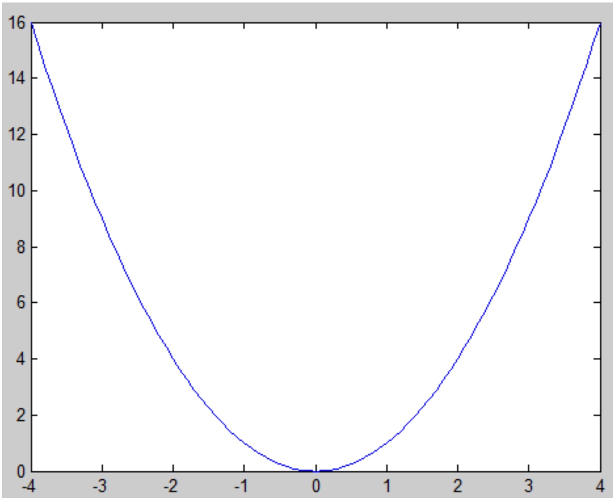


شکل ۷,۱

۸- واضح است که رابطه خطی وجود ندارد.

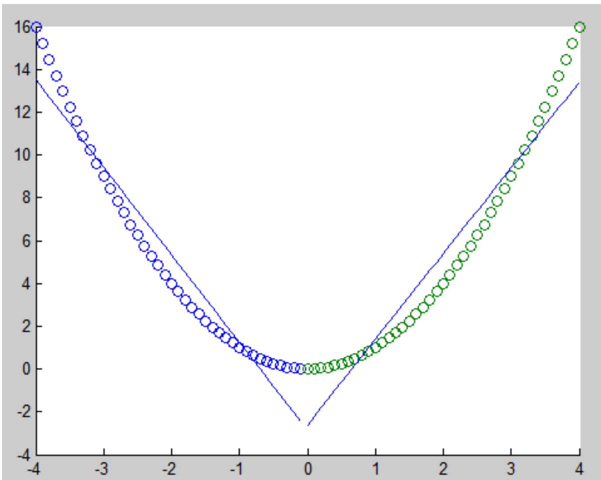
پس از رسم تابع (شکل ۸,۱) دیده می‌شود که می‌توان با دو مدل خطی این تابع را تخمین زد، مدل اول برای قسمت نزولی تابع و مدل دوم برای بخش صعودی آن.(یعنی به صورت محلی) بدین منظور در حالت کلی، نقطه‌ای برای بخش بندی انتخاب می‌شود که معیار هدف را بیشینه کند. در اینجا این معیار می‌تواند مجموع همبستگی خطی دو بخش باشد

.



شکل ۸,۱

بدین ترتیب پس از انتخاب نقطه ی بهینه برای بخش بندی و مدل خطی تخمین زده شده برای هر بخش در شکل زیر (شکل ۸,۲) نمایان است.



شکل ۸,۲