سارا رستمی تمرین ML - سری 2 810100355

1-

الف) روش هیستوگرام - روش ساده‌ای برای بخش‌بندی فضا به تعدادی سلول هم اندازه می‌باشد و سپس محاسبه‌ی هیستوگرام می باشد. روش هیستوگرام برای فهم دیتا مخصوصاً زمانی که داده تک بعدی باشد، بسیار مناسب است.

ب)

مزایای روش parzen :

- می توان از آن برای داده ها از هر توزیعی استفاده کرد.
- در تئوری، می‌توان نشان داد که با رفتن تعداد نمونه‌ها به سمت بی نهایت، همگرا می‌شود.

معایب روش parzen :

- ممکن است به تعداد زیادی داده نیاز داشته باشد تا تخمین دقیقی انجام دهد
- تعداد داده‌های train در عمل محدود است و انتخاب اندازه‌ی پنجره‌ی مناسب (h)، دشوار است
- پیچیدگی محاسباتی بالا (برای classify کردن یک نقطه، باید یک تابع را محاسبه کنیم که احتمالاً به کل نمونه ها بستگی دارد)
- انتخاب سایز پنجره (h) کار ساده‌ای نیست.

مزایای روش KNN :

- می توان از آن برای داده ها از هر نوع توزیعی استفاده کرد (مثل parzen)
- بسیار ساده و قدرتمند است
- اگر اندازه‌ی سمپل به قدر کافی بزرگ باشد، classification را به خوبی انجام می دهد

معایب روش KNN :

- انتخاب بهترین مقدار k، ممکن است دشوار باشد
- همانند parzen پیچیدگی محاسباتی بالایی دارد اما قابل بهبود است
- برای دقت بالا، به تعداد نمونه های زیادی نیاز دارد (مثل parzen)

← هیچ وقت این مشکل را نمی‌توان برطرف کرد مگر با فرض کردن توزیع پارامتریک

ادامه صفحه بعد :

Subject: Date: Sa Su Mo Tu We Th Fr

به طور کلی، روش parzen زمانی خوب جواب می‌دهد که k عدد بزرگی باشد، بنابراین در جاهایی که داده کمی داریم، پاسخ خوبی نمی‌دهد.

بهتر است در مواردی که داده‌ها متراکم هستند از parzen و در مواردی که داده‌ها sparse هستند از knn استفاده کنیم.

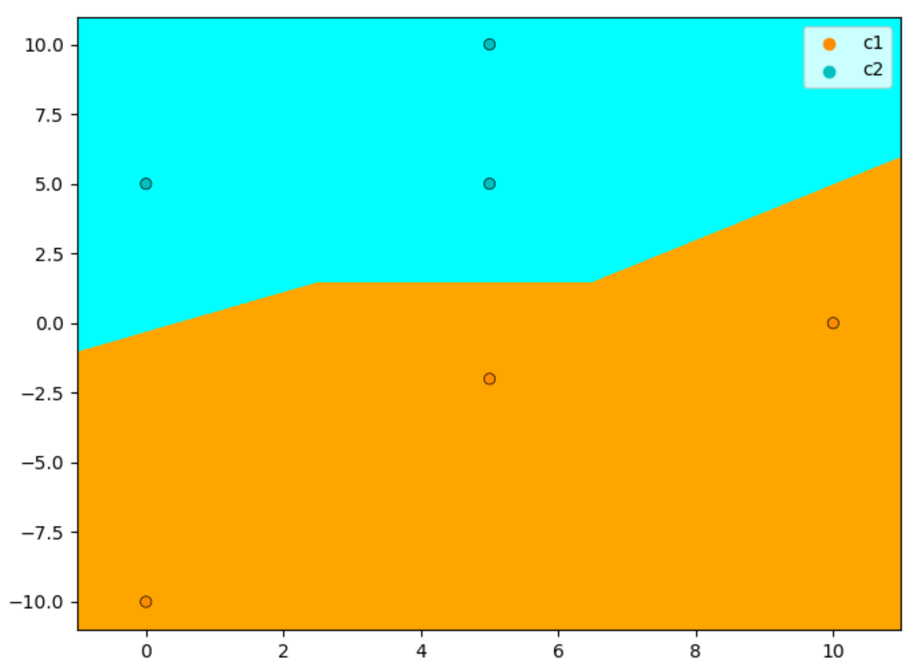
زمانی که pdf یکنواخت باشد، هردو به خوبی کار می‌کنند.

ج)

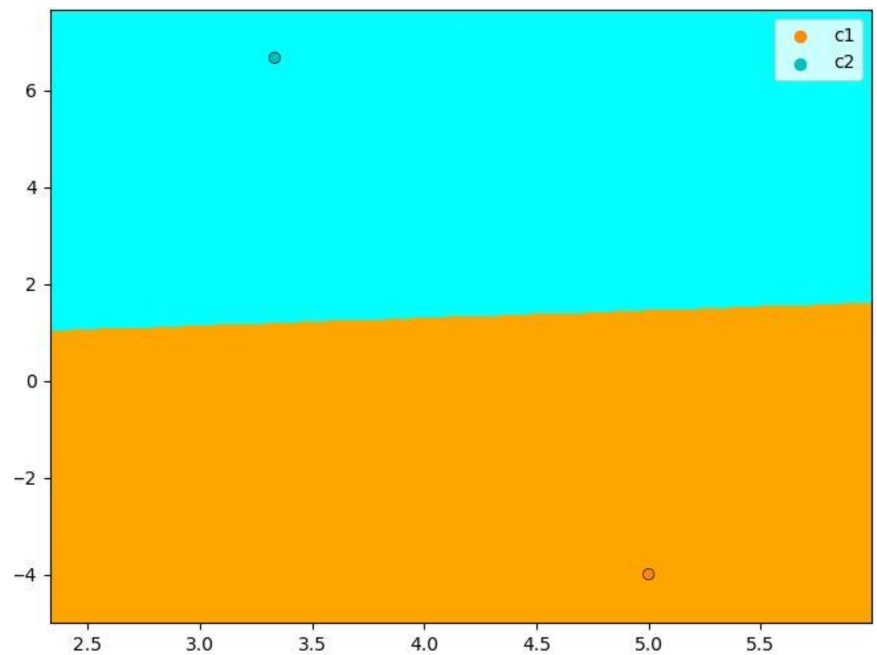
زمانی که احتمال posterior دقیقاً یکسان باشند، ML و MAP یکسان عمل می‌کنند.

۲- با استفاده از کد ضمیمه شده(p2) knn را با k=1 اجرا کردم که نتایج بدست آمده برای هر بخش به صورت زیر است:

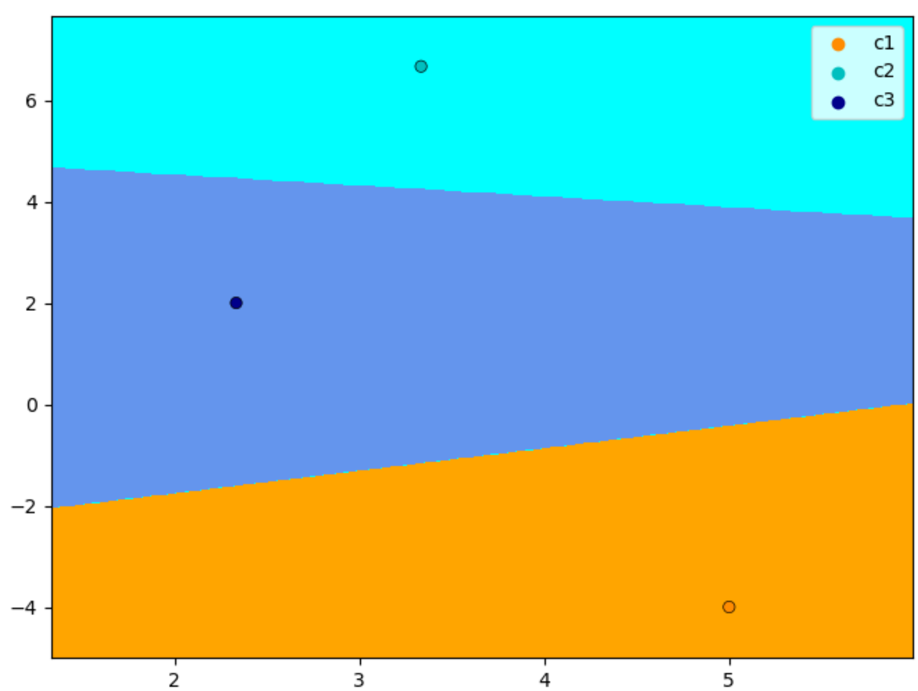
الف)

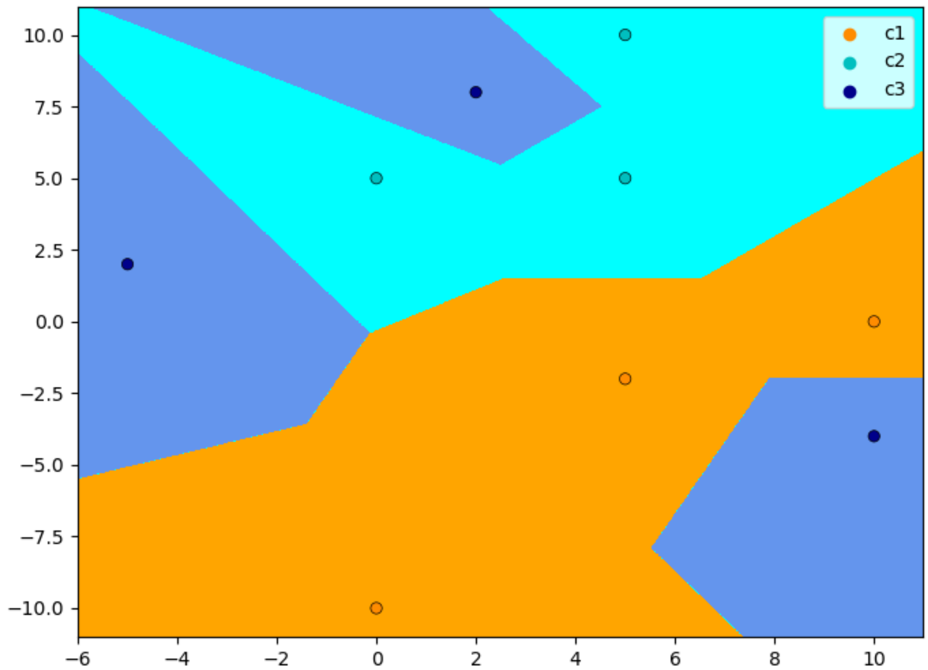


ب)



ج)





c1
c2
c3
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−6
−4
−2
0
2
4
6
8
10

-3

$$P(\omega_i) = \frac{1}{c} \quad , \quad P(x|\omega_i) = \begin{cases} 1 & 0 \leq x \leq \frac{cr}{c-1} \\ 1 & i \leq x \leq i+1 - \frac{cr}{c-1} \\ 0 & \text{otherwise} \end{cases}$$

$$P(x) = \sum_{i=1}^{c} P(x|\omega_i)P(\omega_i) = \begin{cases} 1 & 0 \leq x \leq \frac{cr}{c-1} \\ \frac{1}{c} & i \leq x \leq (i+1) - \frac{cr}{c-1} \quad (*) \\ 0 & \text{otherwise} \end{cases}$$

$$\xrightarrow{(*)} 0 \leq \frac{cr}{c-1} \leq 1 \longrightarrow 0 \leq r \leq \frac{c-1}{c}$$

The Bayes error:

$$① \; P^* = P(e) = \int P(e|x)p(x)\,dx = P(e|0)P(0) + P(e|\text{not } 0)P(\text{not } 0) =$$

$$= \frac{c-1}{c}\,\frac{cr}{c-1} + 0\left(1 - \frac{cr}{c-1}\right) = r$$

The nearest neighbor error rate:

$$② \; P = \int \left[1 - \sum_{i\,1}^{c} P^2(\omega_i|x)\right] p(x)\,dx = \int_0^{cr/(c-1)} \left(1 - \frac{c\left(\frac{1}{c}\right)^2}{p^2(x)}\right) p(x)\,dx \; +$$

$$\sum_{j=1}^{c} \int_j^{j+1-\frac{cr}{c-1}} [1-1]\,p(x)\,dx = \int_0^{\frac{cr}{c-1}} \left(1 - \frac{\frac{1}{c}}{p^2(x)}\right) p(x)\,dx =$$

$$\int_0^{\frac{cr}{c-1}} \left(1 - \frac{1}{c}\right) dx = \left(1 - \frac{1}{c}\right) \frac{cr}{c-1} = r$$

$$\xrightarrow{①,②} P^* = P = r$$

۲-

الف)

$$P(x1) = \begin{cases} \frac{1}{\theta_1} e^{-\theta_1 x_1} & ; x_1 > 0 \\ 0 & ; \text{otherwise} \end{cases}$$

$$P(x2) = U(0, \theta_2) = \begin{cases} \frac{1}{\theta_2} & ; 0 \leq x2 \leq \theta \\ 0 & ; \text{otherwise} \end{cases}$$

$$L(\theta|z) = L(\theta|x,y) = P(x,y|\theta)$$

$\downarrow \log$

$$\log L(\theta|x,y) = \log P(x,y|\theta) = \sum_{i=1}^{n} \log P(x_i, y|\theta) =$$

$$= \sum_{i=1}^{n} \log P_{y_i}(x_i | \theta y_i)$$

$$Q(\theta|\theta^\circ) = E_{Y|X,\theta^\circ}\left[\log L(\theta|x,y)\right] = \int_{-\infty}^{+\infty} \log(P(x_i|\theta))\, P(y|x,\theta)\, dy$$

$$= \int_{-\infty}^{+\infty} \log(P(x_1|\theta)) P(y|X,\theta^\circ) + \log(P(x_2|\theta)) P(y|X,\theta^\circ) +$$

$$\log(P(x_3|\theta)) P(y|X,\theta^\circ) dy = \log(P(x_1|\theta)) \underbrace{\int_{-\infty}^{+\infty} P(y|X,\theta^\circ)\, dy}_{1} +$$

$$\log P(x_2|\theta) \underbrace{\int_{-\infty}^{+\infty} P(y|x,\theta^\circ)\, dy}_{1} + \int_{-\infty}^{+\infty} \log(P(x_3|\theta)) P(y|X,\theta^\circ)\, dy$$

$$= \log P(X^{(1)}|\theta) + \log P(X^{(2)}|\theta) + \int_{-\infty}^{+\infty} \left(\log P(X_1^{(3)}|\theta) + \log P(X_2^{(3)}|\theta)\right) P(X^{(3)}|X,\theta^\circ)\, dx$$

$(*)$

$$\Rightarrow P(X^{(3)}|X,\theta^\circ) = \frac{P(X^{(3)}|\theta^\circ)}{\int_{-\infty}^{+\infty} P(X^{(3)}|\theta^\circ)\, dX^{(3)}}, \quad \theta^\circ = \begin{pmatrix} 2 \\ 4 \end{pmatrix} = \begin{pmatrix} \theta_1^\circ \\ \theta_2^\circ \end{pmatrix}$$

$$P(X^{(3)}|X,\theta) = \frac{P(X_1^{(3)}|\theta^\circ) P(X_2^{(3)}|\theta^\circ)}{\int_{-\infty}^{+\infty} P(X_1^{(3)}|\theta^\circ) P(X_2^{(3)}|\theta^\circ)\, dX_2^{(3)}} = \frac{\cancel{P(X_1^{(3)}|\theta^\circ)} P(X_2^{(3)}|\theta^\circ)}{\cancel{P(X_1^{(3)}|\theta^\circ)} \underbrace{\int_{-\infty}^{+\infty} P(X_2^{(3)}|\theta^\circ)\, dX_2^{(3)}}_{=1}}$$

Subject: Date: Sa Su Mo Tu We Th Fr

ادامه‌ی سؤال ۴، قسمت الف)

(*)

$$\int_{-\infty}^{+\infty} \left(\log P(x_1^{(3)}|\theta) + \log P(x_2^{(3)}|\theta)\right) P(x_2^{(3)}|\theta^\circ)\, dx_2^{(3)}$$

(1) $3 \leq \theta_2 \leq 4$: $\int_0^{\theta_2} \left(\log\left(\frac{1}{\theta_1} e^{-2\theta_1}\right) + \log\left(\frac{1}{\theta_2}\right)\right) \frac{1}{4}\, dx_2^{(3)} =$

$$= \boxed{\frac{\theta_2}{4}\left(-\log\theta_1 - 2\theta_1 - \log\theta_2\right)}$$

(*) $\rightarrow$ = (2) $\theta_2 \geqslant 4$: $\int_0^{4} \left(\log\left(\frac{1}{\theta_1} e^{-2\theta_1}\right) \log\theta_2\right) \frac{1}{4}\, dx_2^{(3)} =$

$$= \boxed{-\log\theta_1 - 2\theta_1 - \log\theta_2}$$

(3) otherwise : $= \boxed{0}$

$$Q(\theta, \theta^\circ) = \log P(x^{(1)}|\theta) + \log P(x^{(2)}|\theta) + (*)$$

$Q(\theta,\theta^\circ)$ سه مقدار می‌تواند داشته باشد:

(1) $\rightarrow Q(\theta,\theta^\circ) = \log\left(P(x_1^{(1)}|\theta) P(x_2^{(1)}|\theta)\right) + \log\left(P(x_1^{(2)}|\theta) P(x_2^{(2)}|\theta)\right)$

$+ \frac{\theta_2}{4}\left(-\log\theta_1 - 2\theta_1 - \log\theta_2\right) = -\log\theta_1 - \theta_1 - \log\theta_2 - \log\theta_1$

$-3\theta_1 - \log\theta_2 + \frac{\theta_2}{4}\left(-\log\theta_1 - 2\theta_1 - \log\theta_2\right) =$

$$\boxed{-2\log\theta_1 - 2\log\theta_2 - 4\theta_1 + \frac{\theta_2}{4}\left(-\log\theta_1 - 2\theta_1 - \log\theta_2\right)}$$

Subject: Date: Sa Su Mo Tu We Th Fr

ادامه‌ی سؤال ۴، قسمت الف)

$$② \rightarrow Q(\theta,\theta^\circ) = \boxed{-3\log\theta_1 - 6\theta_1 - 3\log\theta_2}$$

$$③ \rightarrow Q(\theta,\theta^\circ) = \boxed{-2\log\theta_1 - 2\log\theta_2 - 3\theta_1}$$

ب)

$$P(x_1) = \frac{1}{\theta_1} e^{-\theta_1 x_1}$$

$$\int_{-\infty}^{+\infty} P(x_1)\,dx_1 = 1 \rightarrow \int_{-\infty}^{+\infty} \frac{1}{\theta_1} e^{-\theta_1 x_1} dx_1 = \int_{0}^{+\infty} \frac{1}{\theta_1} e^{-\theta_1 x_1} dx_1$$

$$= \frac{1}{\theta_1}\left[-\frac{1}{\theta_1} e^{-\theta_1 x_1}\right|_0^{+\infty} = \frac{1}{\theta_1}\left[0 + \frac{1}{\theta_1}\right] = 1 \rightarrow \frac{1}{(\theta_1)^2} = 1$$

همواره $\theta_1 = 1$

$(3 \leq \theta_2 \leq 4)$

$$\xrightarrow[\text{از قسمت الف بدست آوردیم}]{\text{حالت ①}} Q(\theta,\theta^\circ) = -2\log\theta_2 - 4 - \frac{\theta_2}{2} - \frac{\theta_2}{4}\log\theta_2$$

به ازای $\theta_2 = 3$ ، $Q(\theta,\theta^\circ)$ ماکسیمم می‌شود.

$$\xrightarrow{\theta_2=3} = -2\underbrace{\log(3)}_{0.477} - 4 - \frac{3}{2} - \frac{3}{4}\log(3) = -6.81$$

$$\xrightarrow[(\theta_2 \geq 4)]{\text{حالت ②}} Q(\theta,\theta^\circ) = -3\log\theta_1 - 6\theta_1 - 3\log\theta_2$$

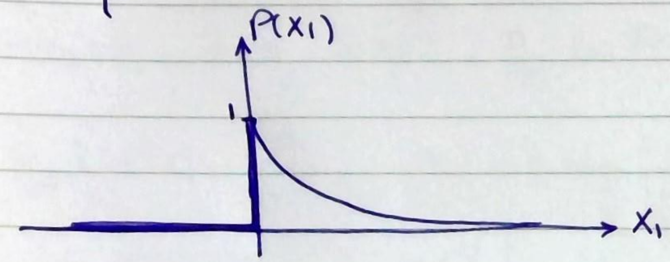
به ازای $\theta_2 = 4$ ، $Q(\theta,\theta^\circ)$ در این حالت ماکسیمم می‌شود.

$$\xrightarrow{\theta_2=4} = -3\times 0 - 6 - 3\log 4 = -12$$

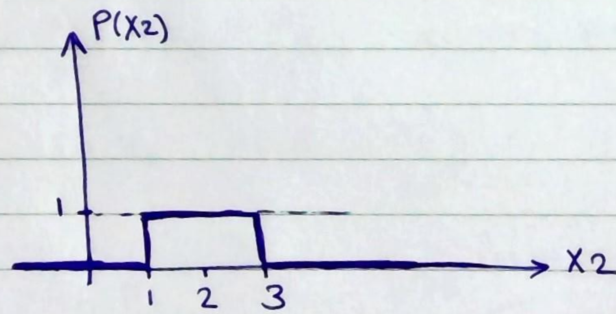
$$\left.\boxed{\begin{matrix}\theta_2 = 3 \\ \theta_1 = 1\end{matrix}}\right\}$$ مقدار جدید برای θ ها $\Longleftarrow$ حالت ② < حالت ①

PAPCO

$$P(X_1) = \begin{cases} \frac{1}{\theta_1} e^{-\theta_1 x_1} & ; x > 0 \\ 0 & ; \text{otherwise} \end{cases} \xrightarrow{\theta_1 = 1} P(x_1) = \begin{cases} e^{-x_1} & ; x > 0 \\ 0 & ; \text{otherwise} \end{cases}$$



$$P(X_2) = \begin{cases} \frac{1}{\theta_2} & ; 0 < x_2 < \theta \\ 0 & ; \text{otherwise} \end{cases} \xrightarrow{\theta_2 = 3} P(X_2) = \begin{cases} \frac{1}{3} & ; 0 < x < 3 \\ 0 & ; \text{otherwise} \end{cases}$$



Subject: Date: Sa Su Mo Tu We Th Fr

_5

1. $f(x_k;\theta) = \theta \exp(-\theta x_k) \quad ; \quad x_k \geqslant 0 \ , \ \theta > 0$

$$l(\theta|X) = \sum_{i=1}^{n} \ln\left(P(x_k|\theta)\right) \longrightarrow \frac{\partial}{\partial\theta} l(\theta|X) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln P(x_k|\theta)$$

$$\ln P(x_k|\theta) = \ln\theta - \theta x_k \longrightarrow \frac{\partial}{\partial\theta} \ln P(x_k|\theta) = \frac{1}{\theta} - x_k$$

$$\Rightarrow \frac{\partial}{\partial\theta} l(\theta|X) = \sum_{i=1}^{n} \frac{1}{\hat\theta} - x_k = 0 \Longrightarrow \frac{n}{\hat\theta} - \sum_{i=1}^{n} x_k = 0$$

$$\Longrightarrow \hat\theta = \frac{n}{\sum_{i=1}^{n} x_k} \Longrightarrow \hat\theta = \frac{1}{x}$$

2. $f(x_k;\theta) = \frac{x_k}{\theta^2} \exp\left(-\frac{x^2{}_k}{2\theta^2}\right) \ ; \ x_k \geqslant 0 \ , \ \theta > 0$

$$\ln P(x_k|\theta) = \ln x_k - 2\ln\theta - \frac{x_k^2}{2\theta^2}$$

$$\frac{\partial}{\partial\theta} \ln P(x_k|\theta) = \frac{-2}{\theta} + \frac{x_k^2}{\theta^3} \Rightarrow \frac{\partial}{\partial\theta} \ln(\theta|X) = \sum_{k=1}^{n} \frac{x_k^2}{\hat\theta^3} - \frac{2}{\hat\theta} = 0$$

$$\frac{\sum_{k=1}^{n} x_k^2}{\hat\theta^3} = \frac{2n}{\theta} \Rightarrow \qquad \hat\theta = \sqrt{\frac{\sum_{k=1}^{n} x_k^2}{2n}}$$

3. $f(x_k;\theta) = \sqrt{\theta}\, x_k^{\sqrt{\theta-1}} \quad ; \quad 0 \leq x_k \leq 1 \ , \ \theta > 0$

$$\ln P(x_k|\theta) = \frac{1}{2}\ln\theta + \sqrt{\theta-1}\ \ln x_k \Rightarrow \frac{\partial}{\partial\theta} \ln P(x_k|\theta) = \frac{1}{2\theta} +$$

$$\frac{1}{2\sqrt{\theta-1}} \ln(x_k) \qquad \frac{\partial}{\partial\theta} \ln\theta(x) = \sum_{k=1}^{n} \frac{1}{2\hat\theta} + \frac{1}{2\sqrt{\hat\theta-1}} \ln(x_k) = 0$$

$$\frac{-n}{\hat\theta} = \frac{\sum_{k=1}^{n} \ln x_k}{\sqrt{\hat\theta-1}} \Rightarrow \frac{\sqrt{-\hat\theta-1}}{\hat\theta} = \frac{\sum_{k=1}^{n} \ln x_k}{n} \longrightarrow \frac{\theta-1}{\theta^2} = \alpha^2$$

Subject: Date: Sa Su Mo Tu We Th Fr

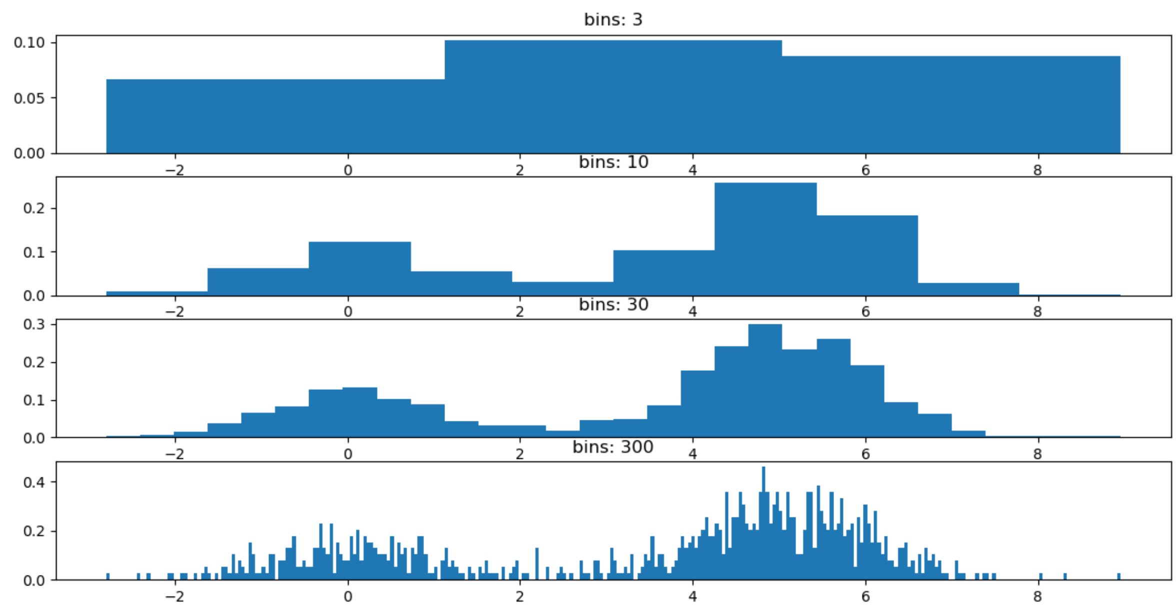
ادامه سؤال 5، قسمت 3)

$$\alpha^2\theta^2 = \theta - 1 \Rightarrow \alpha^2\theta^2 - \theta + 1 = 0$$

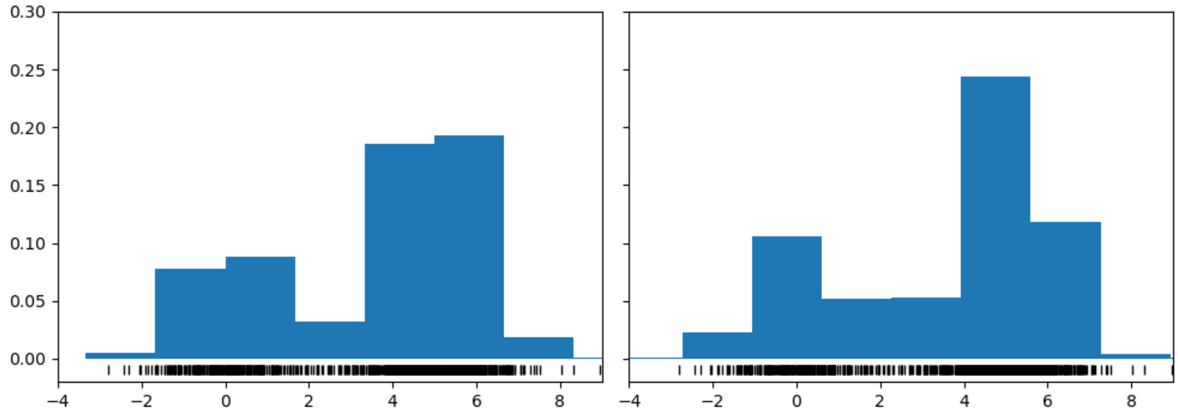
$$\Rightarrow \theta = \frac{1 \pm \sqrt{1-4\alpha^2}}{2\alpha^2}$$

۶-

الف) با بزرگ شدن تعداد binها و درنتیجه‌ی آن کوچک شدن bin width نمودار هیستوگرام(تخمین تابع چگالی) spiky تر می‌شود. اما زمانی که تعداد binها کم است(یعنی زمانی که bin width بزرگ است)، نمودار smoothتر می‌باشد چرا که averaging بزرگی روی تعداد زیادی از داده‌ها داریم انجام می‌دیم.



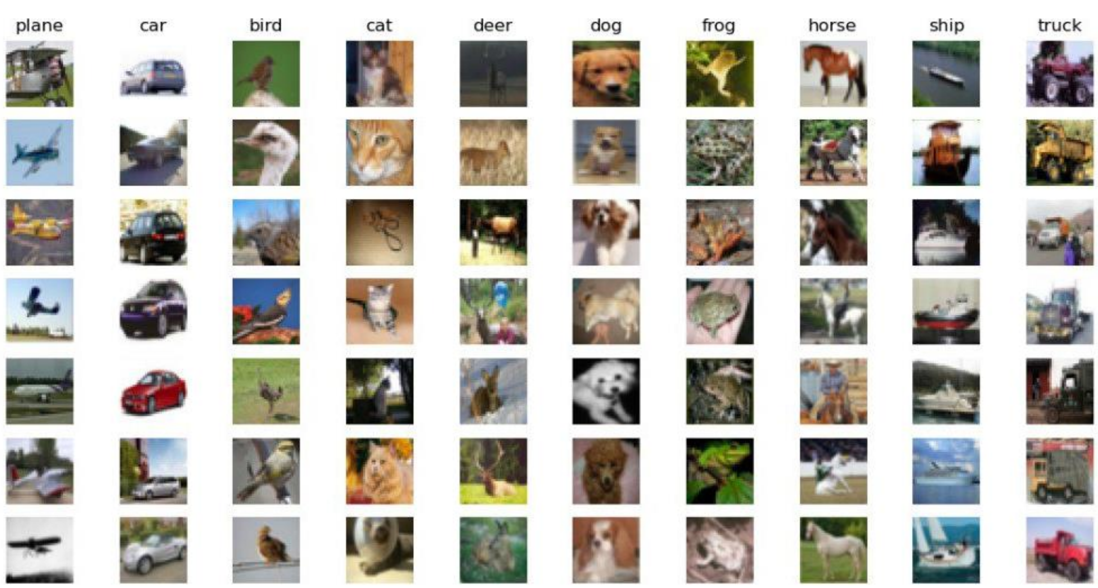
ب)



یکی از مشکلات روش هیستوگرام برای تخمین چگالی این است که انتخاب سایز bin های متفاوت و locationمتفاوت برای bin ها می‌تواند منجر به representationهای متفاوتی از داده شود که ویژگی‌های متفاوتی دارند. یکی از دلایل تفاوت در تخمین چگالی‌های دو شکل بالا این است که ارتفاع bin اغلب نشان‌دهنده‌ی چگالی داده‌های اطراف آن bin نمی‌باشد بلکه نشان‌دهنده‌ی این است که داده‌ها در چه binهای قرار می‌گیرند و در واقع ارتفاع هر bin به تعداد داده‌هایی که به آن bin منطبق می‌شوند بستگی دارد. همچنین دلیل دیگر تفاوت در دو شکل بالا این است که مرز binها در این دو شکل با هم متفاوت است.

۸- قسمت‌هایی که در گزارش نیامده در کد مربوط به سوال که پیوست شده قرار دادم.

الف)



د) انتظار میرود که فاصله منهتنی مقیاس مناسبتری نسبت به فاصله اقلیدسی برای تعیین فاصله‌ی دو تصویر باشد. فاصله‌ی اقلیدسی توجه زیادی به outlierها می کند در حالیکه فاصله منهتنی اهمیت یکسانی برای تمام نقاط قائل می‌شود. به طور مثال فرض کنید که رنگ پس زمینه دو تصویر گربه بسیار متفاوت باشد. طبق معیار فاصله‌ی اقلیدسی اختلاف این دو تصویر مقدار خیلی زیادی پیدا می‌کند که مطلوب نمی‌باشد. نتایج حاصل بر اساس معیار f1_score با اجرای knn به ازای مقادیر k = 3, 7, 15 به ازای هر کدام از فواصل به شرح زیر است:

weighted f1 where euclidean, k=03 : 0.3191

weighted f1 where manhattan, k=03: 0.3596

weighted f1 where euclidean, k=07: 0.3208

weighted f1 where manhattan, k=07: 0.3702 <-- best result

weighted f1 where euclidean, k=15: 0.3231

weighted f1 where manhattan, k=15: 0.3662

می بینیم که نتایج بدست آمده با پیش‌بینی ما همخوانی دارد و به ازای هر مقدار k میزان f1_score با فاصله منهتنی بیشتر از میزان f1_score با فاصله اقلیدسی است.

همچنین می‌بینیم که به ازای k=7 طبقه‌بند بهتر عمل می‌کند.