به نام خدا

دانشگاه تهران

دانشکده مهندسی برق و کامپیوتر

درس یادگیری ماشین

تمرین دوم

نام و نام خانوادگی	امیرحسین پورداود
شماره دانشجویی	۸۱۰۱۰۱۱۲۰
تاریخ ارسال گزارش	۱۴۰۲.۰۹.۰۳

فهرست

پاسخ ۱.

۱-۱. قسمت ۱

روش پارزن:

در روش پارزن h_n نشانگر بزرگی همسایگی در نظر گرفته شده است. اگر h_n کوچک باشد، pdf تخمینی دارای تغییرات زیاد و بصورت شارپی خواهد بود، اگر h_n بزرگ باشد، pdf تخمینی خیلی نرم خواهد بود و تغییرات pdf را نشان نخواهد داد.

روش KNN:

در روش KNN، k_n نشان دهنده تعداد نمونه های داخل همسایگی است. در این مورد نیز همانند روش پارزن اگر k_n کوچک باشد، pdf بصورت شارپی و اگر بزرگ باشد، pdf خیلی نرم خواهد شد.

برای مشخص کردن k_n و h_n باید با توجه به تعداد نمونه های در دست، تصمیم گرفت.

۱-۲. قسمت ۲

برای این قسمت باید شرط $\int p(x)dx = 1$ را بررسی کنیم، حال در نقطه‌ای همانند x_i احتمال برابر $p(x_i) = \frac{K}{NV_i}$ است.

با فرض اینکه V_i ها کوچک اند می توان انتگرال فوق را بصورت زیر بازنویسی کرد

$$\int p(x)\,dx = \sum_{i=1}^{N} p(x_i)\,dx_i = \sum_{i=1}^{N} \frac{K}{NV_i}\cdot V_i = \frac{K}{N}\sum_{i=1}^{N} 1 = K$$

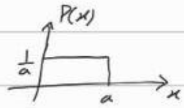
درصورت $K \neq 1$ این انتگرال شرط مورد نظر را ارضا نمی‌کند، بنابراین واگرا می شود.

پاسخ ۲.

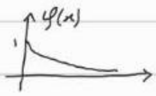
۲-۱. قسمت ۱

Q2/

$$p(x) \sim U(0, a) \Rightarrow$$



$$\varphi(x) = \begin{cases} e^{-x} & ; x > 0 \\ 0 & ; x \leq 0 \end{cases} \Rightarrow$$

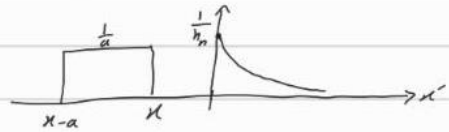


① $$\bar{P}_n(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{a}(1 - e^{-\frac{x}{h_n}}) & 0 \leq x \leq a \\ \frac{1}{a}(e^{\frac{a}{h_n}} - 1) e^{-\frac{x}{h_n}} & x \geq a \end{cases}$$

$$\bar{P}_n(x) = \varphi_{h_n}(x) * p(x) \qquad \varphi_{h_n}(x) = \frac{1}{h_n}\varphi\left(\frac{x}{h_n}\right) = \begin{cases} \frac{1}{h_n} e^{-\frac{x}{h_n}} & x > 0 \\ 0 & o.w. \end{cases}$$

$$\Rightarrow P_n(x) = \int \frac{1}{h_n}\varphi\left(\frac{x'}{h_n}\right) p(x - x')\, dx'$$

$\Rightarrow$



if $x < 0 \Rightarrow P_n(x) = 0$

if $0 \leq x \leq a \Rightarrow \int_0^x \frac{1}{h_n} e^{-\frac{x'}{h_n}} \cdot \frac{1}{a}\, dx' = \frac{1}{a}(1 - e^{-\frac{x}{h_n}})$

if $x > a \Rightarrow \int_{x-a}^{x} \frac{1}{h_n} e^{-\frac{x'}{h_n}} \cdot \frac{1}{a}\, dx' = \frac{-1}{a} e^{-\frac{x'}{h_n}}\Big]_{x-a}^{x} = \frac{1}{a}(e^{\frac{a}{h_n}} - 1)e^{-\frac{x}{h_n}}$

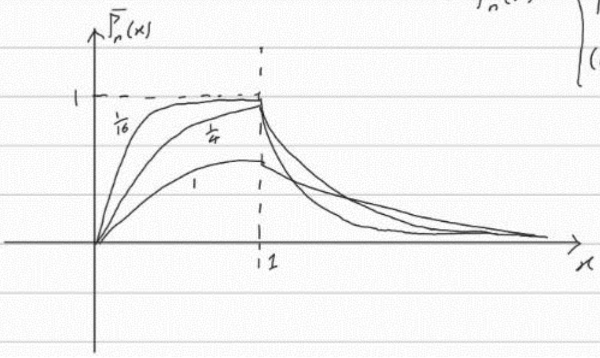
☑

۲-۲. قسمت ۲

2

$a=1 \quad , \quad h_n=\left\{1, \frac{1}{4}, \frac{1}{16}\right\}$

$$\Rightarrow \bar{P}_n(x)=\begin{cases} 0 & x<0 \\ 1-e^{-\frac{x}{h_n}} & 0\leq x\leq 1 \\ (e^{\frac{1}{h_n}}-1)\,e^{-\frac{x}{h_n}} & x\geq 1 \end{cases}$$



۲-۳. قسمت ۳

3

$\left|\frac{\bar{P}_n(x)-P(x)}{P(x)}\right|\leq 0.01$ for $0<x<a$ $\qquad \bar{P}_n(x)=\frac{1}{a}(1-e^{-\frac{x}{h_n}})$

$P(x)=\frac{1}{a}$

$$\Rightarrow \left|\frac{\cancel{\frac{1}{a}}(1-e^{-\frac{x}{h_n}})-\cancel{\frac{1}{a}}}{\cancel{\frac{1}{a}}}\right| = e^{\frac{-x}{h_n}}\leq 0.01 \xrightarrow{\ln} \frac{-x}{h_n}=\ln^{0.01}=-\ln^{100}$$

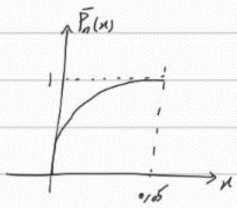
$$\Rightarrow h_n=\frac{x}{\ln^{100}}$$

$$x=\frac{a}{100} \Rightarrow h_n=\frac{a}{100\ln^{100}}$$

۲-۴. قسمت ۴

4 $\quad a=1$

$0\leq x\leq 0.05 \longrightarrow \bar{P}_n(x)=1-e^{-\frac{x}{h_n}} \quad , \quad h_n=\frac{a}{100\ln^{100}}=\frac{1}{100\ln^{100}}=22\times10^{-4}$



پاسخ ۳.

۳-۱. قسمت ۱

Q3/

$p(x) \sim N(\mu, \sigma^2)$, $\varphi(x) = N(0,1)$

$$p_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} \varphi\left(\frac{x - x_i}{h_n}\right)$$

①

$\bar{p}_n(x) \sim N(\mu, h_n^2 + \sigma^2)$

$$\Rightarrow \bar{p}_n(x) = \varphi_{h_n}(x) * p(x) = \underbrace{\frac{1}{h_n}\varphi\left(\frac{x}{h_n}\right)}_{N(0, h_n^2)} * \underbrace{p(x)}_{N(\mu, \sigma^2)}$$

$$p(x) \sim N(\mu, \sigma^2) \xrightarrow{MGF} e^{t\mu + \frac{1}{2}\sigma^2 t^2}$$

$$\varphi\left(\frac{x}{h_n}\right) \sim N(0, h_n^2) \xrightarrow{\text{"}} e^{\frac{1}{2} h_n^2 t^2}$$

$$\Rightarrow MGF(\bar{p}_n(x)) = MGF\left(\frac{1}{h_n}\varphi\left(\frac{x}{h_n}\right)\right) \cdot MGF(p(x))$$

$$= e^{\frac{1}{2} h_n^2 t^2} \cdot e^{t\mu + \frac{1}{2}\sigma^2 t^2} = e^{t\mu + \frac{1}{2}(h_n^2 + \sigma^2)t^2}$$

$$\Rightarrow \bar{p}_n(x) \sim N(\mu, \sigma^2 + h_n^2) \quad \boxed{\checkmark}$$

۳-۲. قسمت ۲

(3)

$$p(x) - \bar{p}_n(x) \cong \frac{1}{2}\left(\frac{h_n}{\sigma}\right)^2\left[1 - \frac{(x-\mu)^2}{\sigma^2}\right]p(x)$$

$$p(x) - \bar{p}_n(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) - \frac{1}{\sqrt{2\pi(\sigma^2+h_n^2)}}\exp\left(-\frac{(x-\mu)^2}{2(\sigma^2+h_n^2)}\right)$$

$$= p(x)\left[1 - \underbrace{\sqrt{\frac{\cancel{2\pi}\sigma^2}{\cancel{2\pi}(\sigma^2+h_n^2)}}}_{\sqrt{\frac{1}{1+\left(\frac{h_n}{\sigma}\right)^2}} \overset{\frac{h_n}{\sigma}\to 0}{\sim} 1-\frac{1}{2}\left(\frac{h_n}{\sigma}\right)^2}\exp\left(-\frac{(x-\mu)^2}{2}\underbrace{\left[\frac{1}{\sigma^2+h_n^2}-\frac{1}{\sigma^2}\right]}_{\frac{-h_n^2}{\sigma^2(\sigma^2+h_n^2)}}\right)\right]$$

$$e^x \overset{x\to 0}{\sim} 1+x$$

$$= p(x)\left[1 - \left(1-\frac{1}{2}\left(\frac{h_n}{\sigma}\right)^2\right)\left(1+\frac{h_n^2}{2\sigma^2}\cdot\frac{(x-\mu)^2}{\sigma^2+h_n^2}\right)\right]$$

$$= p(x)\left[\cancel{1}-\cancel{1}+\frac{1}{2}\frac{h_n^2}{\sigma^2} - \frac{h_n^2}{2\sigma^2}\,\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right]$$

$$= p(x)\left[1-\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right]\frac{1}{2}\frac{h_n^2}{\sigma^2}$$

۳-۳. قسمت ۳

(3)

$$Var[P_n(x)] \simeq \frac{1}{2n h_n \sqrt{\pi}} P(x)$$

$$Var[P_n(x)] = Var\left[\frac{1}{n h_n} \sum_{i=1}^{n} \varphi\left(\frac{x-x_i}{h_n}\right)\right] = \frac{1}{n^2 h_n^2} \sum_{i=1}^{n} Var\left[\varphi\left(\frac{x-x_i}{h_n}\right)\right]$$

$$= \frac{1}{n^{\cancel{2}} h_n^2} \cancel{n} \times Var\left[\varphi\left(\frac{x-x_i}{h_n}\right)\right]$$

$$Var\left[\varphi\left(\frac{x-x_i}{h_n}\right)\right] = E\left[\varphi^2\left(\frac{x-x_i}{h_n}\right)\right] - E^2\left[\varphi\left(\frac{x-x_i}{h_n}\right)\right]$$

$$E\left[\varphi^2\left(\frac{x-x_i}{h_n}\right)\right] = \int \varphi^2\left(\frac{x-x'}{h_n}\right) P(x')\,dx' = \int \frac{1}{2\pi} e^{-\frac{(x-x')^2}{h_n^2}} \cdot e^{-\frac{(x'-\mu)^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi\sigma^2}}\,dx'$$

$$= \frac{1}{\sqrt{2\pi}} \int \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-x'}{\frac{h_n}{\sqrt{2}}}\right)^2} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x'-\mu)^2}{2\sigma^2}}\,dx'$$

$$= \frac{\frac{h_n}{\sqrt{2}}}{\sqrt{2\pi}} N\left(0, \frac{h_n^2}{2}\right) * N(\mu, \sigma^2) \qquad \text{from (a)}$$

$$= \frac{\frac{h_n}{\sqrt{2}}}{\sqrt{2\pi}} N\left(\mu, \sigma^2 + \frac{h_n^2}{2}\right)$$

$$\frac{1}{n h_n^2} E\left\{\varphi^2\left(\frac{x-x'}{h_n}\right)\right\} = \frac{1}{n h_n^{\cancel{2}}} \cdot \frac{\cancel{h_n}}{\sqrt{2}} \cdot \underbrace{\frac{1}{2\pi \underbrace{\sqrt{\sigma^2 + \frac{h_n}{2}}}_{\sim \sigma}} e^{-\frac{(x-\mu)^2}{\underbrace{2\left(\sigma^2 + \frac{h_n^2}{2}\right)}_{\sim \sigma^2}}}}_{P(x)} = \frac{1}{2n h_n \sqrt{\pi}} P(x) \quad \checkmark$$

$$E\left[\varphi\left(\frac{x-x'}{h_n}\right)\right] = \left[h_n \int \frac{1}{\sqrt{2\pi} h_n} e^{-\frac{(x-x')^2}{2h_n^2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(x'-\mu)^2}{2\sigma^2}}\,dx'\right]$$

$$= \left[h_n N(0, h_n^2) * N(\mu, \sigma^2)\right]^2 = h_n N(\mu, \sigma^2 + h_n^2) \simeq 0 \quad \text{for small } h_n \simeq 0$$

$$Var[P_n(x)] = \frac{1}{2n h_n \sqrt{\pi}} P(x) - 0^2 = \boxed{\frac{1}{2n h_n \sqrt{\pi}}}$$

پاسخ ۴.

$$P(\theta|x) \propto P(x|\theta)P(\theta) \propto \left(\sum_z Q(z)\frac{P(x,z|\theta)}{Q(z)}\right)P(\theta)$$

$$\Rightarrow \ln\left(\sum_z Q(z)\frac{P(x,z|\theta)}{Q(z)}\right) \geqslant \sum_z Q(z)\left[\ln P(x,z|\theta) - \ln Q(z)\right]$$

۴-۱. قسمت ۱

(1)

$$l = P(x|\theta)P(\theta) = \left(\sum_z Q(z)\frac{P(x,z|\theta)}{Q(z)}\right)P(\theta)$$

$$L = \ln\left(\sum_z Q(z)\frac{P(x,z|\theta)}{Q(z)}\right) + \ln P(\theta)$$

Jensons $\Rightarrow \ln\left(\sum_z Q(z)\frac{P(x,z|\theta)}{Q(z)}\right) + \ln P(\theta) \geqslant \sum_z Q(z)\left[\ln P(x,z|\theta) - \ln Q(z)\right] + \ln P(\theta)$

$$\Rightarrow L = \sum_z Q(z)\left[\ln P(x,z|\theta) - \ln Q(z)\right] + \ln P(\theta)$$

E-step:

$$\frac{\partial L}{\partial Q(i)} = \ln P(x,z|\theta) - 1 - \ln Q(i) = 0 \Rightarrow Q(i) = c\,P(z=i,x|\theta)$$

$$\Rightarrow \sum_i Q(i) = 1 \Rightarrow \sum_i c\,P(z=i,x|\theta) = 1$$

$$\Rightarrow c = \frac{1}{\sum_i P(z=i,x|\theta)} \checkmark$$

$$\Rightarrow Q(z) = \frac{P(z,x|\theta)}{\sum_{i=1}^{k} P(z=i,x|\theta)} = \frac{P(z,x|\theta)}{P(x|\theta)} = P(z|x,\theta) \boxed{\checkmark}$$

M-step:

$$\max_\theta L = \max \sum_z Q(z)\ln P(z,x|\theta) - \cancel{Q(z)\ln Q(z)} + \ln P(\theta)$$

$$= \max \sum_z Q(z)\ln P(z,x|\theta) + \ln P(\theta)$$

$Q(z) = P(z|x,\theta^{old})$

$$\Rightarrow \max_\theta \sum_z P(z|x,\theta^{old})\ln(z,x|\theta) + \ln P(\theta) \boxed{\checkmark}$$

۴-۲. قسمت ۲

(3)

$X \sim$

state	Prob.	Num.	
A	$\frac{1}{3}$	n_a	$\rightarrow$ hidden
B	$\frac{1}{3}(1-\theta)$	n_b	$\rightarrow$ visible
C	$\frac{2}{3}\theta$	n_c	$\rightarrow$ hidden
D	$\frac{1}{3}(1-\theta)$	n_d	$\rightarrow$ visible

$$P(\theta) = \frac{\Gamma(\nu_1 + \nu_2)}{\Gamma(\nu_1)\,\Gamma(\nu_2)}\,\theta^{\nu_1 - 1}(1-\theta)^{\nu_2 - 1}$$

E-step:

$$\text{known} \Rightarrow \begin{cases} P(B \mid x, \theta) = 1 \text{ or } 0 \\ P(D \mid x, \theta) = 1 \text{ or } 0 \end{cases}$$

با توجه به داده ها چون B و D را می‌دانیم پس احتمال آنها ۱ یا ۰ است.

$$\text{hidden} \Rightarrow \begin{cases} P(A \mid x, \theta) = \dfrac{P(A, x \mid \theta)}{\sum_{i \in \{A, C\}} P(z=i, x \mid \theta)} = \dfrac{\frac{1}{3}}{\frac{1}{3} + \frac{2}{3}\theta} = \dfrac{1}{1+2\theta} \\ P(C \mid x, \theta) = \dfrac{P(C, x \mid \theta)}{\sum_{i \in \{A, C\}} P(z=i, x \mid \theta)} = \dfrac{\frac{2}{3}\theta}{\frac{1}{3} + \frac{2}{3}\theta} = \dfrac{2\theta}{1+2\theta} \end{cases}$$

$$\beta = n_a + n_c = N - n_b - n_d$$

M-step:

$$\arg\max_{\theta} L(\theta)$$

$$L = \sum_{N} \sum_{Z} P(z|\theta^{old}, x_n) \ln P(x,z|\theta) + \ln P(\theta)$$

$$= n_b \ln\left(\tfrac{1}{3}(1-\theta)\right) + n_d \ln\left(\tfrac{1}{3}(1-\theta)\right)$$

$$+ \sum_{N} \sum_{z\in\{A,C\}} P(z|\theta^{old}, x_n) \ln P(x,z|\theta) + \ln P(\theta)$$

$$= n_b \ln \tfrac{1}{3}(1-\theta) + n_d \ln \tfrac{1}{3}(1-\theta) + \sum_{N} P(A|\theta,x) \ln P(A,x|\theta) + P(C|\theta,x) \ln P(C,x|\theta)$$

$$= n_b \ln \tfrac{1}{3}(1-\theta) + n_d \ln \tfrac{1}{3}(1-\theta) + \beta \left[\frac{1}{1+2\theta^{old}} \ln \frac{1}{3} + \frac{2\theta^{old}}{1+2\theta^{old}} \ln \frac{2\theta}{3} \right] + \ln P(\theta)$$

$$\frac{\partial L}{\partial \theta} = (n_b + n_d) \frac{-\cancel{\tfrac{1}{3}}}{\cancel{\tfrac{1}{3}}(1-\theta)} + \beta \frac{2\theta^{old}}{1+2\theta^{old}} \cdot \frac{\cancel{\tfrac{2}{3}}}{\cancel{\tfrac{2}{3}}\theta} + \underbrace{\frac{1}{P(\theta)} \cdot \frac{\partial P(\theta)}{\partial \theta}}_{\frac{\nu_1-1}{\theta} + \frac{\nu_2-1}{1-\theta}} = 0$$

$$\frac{\partial P(\theta)}{\partial \theta} = \frac{\partial}{\partial \theta} A\, \theta^{\nu_1-1}(1-\theta)^{\nu_2-1} = A(\nu_1-1) \frac{\theta^{\nu_1-1}}{\theta} (1-\theta)^{\nu_2-1} + A(\nu_2-1)\, \theta^{\nu_1-1} \frac{(1-\theta)^{\nu_2-1}}{1-\theta}$$

$$= (\nu_1-1) \frac{P(\theta)}{\theta} + (\nu_2-1) \frac{P(\theta)}{1-\theta}$$

$$\Rightarrow \frac{\partial L}{\partial \theta} = \frac{-(n_b+n_d) + \nu_2 - 1}{1-\theta} + \frac{\beta \frac{2\theta^{old}}{1+2\theta^{old}} + \nu_1 - 1}{\theta} = 0$$

$$\theta = \frac{\beta \frac{2\theta^{old}}{1+2\theta^{old}} + \nu_1 - 1}{\beta \frac{2\theta^{old}}{1+2\theta^{old}} + \nu_1 \cancel{-1} + (n_b+n_d) - \nu_2 \cancel{+1}} \qquad \boxed{\checkmark}$$

پاسخ ۵.

Known → تعداد بردها , تعداد بازی‌ها , تعداد کل افراد

hidden → بازیکن k در روز t

احتمال برد بازیکن k ~ P_k

۵-۱. قسمت ۱

①

$$\text{I)} \quad p(w_t, m_t \mid k) = \binom{m_t}{w_t} P_k^{w_t} (1-P_k)^{m_t-w_t} \qquad \forall t \in 1, \dots, N$$

$$\text{II)} \quad p(w_t, m_t \mid C_t) = \binom{m_t}{w_t} P_{C_t}^{w_t} (1-P_{C_t})^{m_t-w_t}$$

۵-۲. قسمت ۲

② E-step

$$Q(k) = p(k \mid x, \theta) \propto p(x \mid C_t = k, \theta)\, p(C_t = k) \propto \binom{m_t}{x} P_k^{x} (1-P_k)^{m_t-x} \cdot \pi_k$$

$$Q(k) = C \binom{m_t}{x} P_k^{x} (1-P_k)^{m_t-x} \cdot \pi_k$$

$$\sum_k Q(k) = 1 \Rightarrow C = \frac{1}{\sum_k \binom{m_t}{x} P_k^{x} (1-P_k)^{m_t-x} \cdot \pi_k}$$

$$\Rightarrow Q_t^i[k] = \frac{\binom{m_t}{x} P_k^{x} (1-P_k)^{m_t-x} \cdot \pi_k}{\sum_k \binom{m_t}{x} P_k^{x} (1-P_k)^{m_t-x} \cdot \pi_k} \quad ☑$$

۵-۳. قسمت ۳

(3) M-step

$$\pi^i[k] = \frac{\sum_{t=1}^{N} Q_t^i[k]}{n}$$

$$L = \sum_N \sum_k Q_t^i[k] \log P(w_t \mid P_k) = \sum_N \sum_K Q_t^i[k] \log \binom{m_t}{w_t} P_k^{w_t} (1-P_k)^{m_t - w_t}$$

$$= \sum_N \sum_K Q_t^i[k] \left[\log \binom{m_t}{w_t} + w_t \log P_k + (m_t - w_t) \log (1 - P_k) \right]$$

$$\frac{\partial L}{\partial P_j} = \sum_N Q_t^i[j] \left[\frac{w_t}{P_j} + \frac{w_t - m_t}{1 - P_j} \right] = 0$$

$$= \sum_N Q_t^i[j] \frac{(1-P_j) w_t + P_j (w_t - m_t)}{P_j (1 - P_j)} = 0 \Rightarrow \sum_N Q_t^i[j] (-m_t P_j + w_t) = 0$$

$$\Rightarrow P_j \sum_N Q_t^i[j] m_t = \sum_N Q_t^j[j] w_t$$

$$\Rightarrow P_j = \frac{\sum_{t=1}^{N} Q_t^j[j] w_t}{\sum_{t=1}^{N} Q_t^i[j] m_t}$$

پاسخ ۶. شبیه سازی KNN

۶-۱. قسمت ۱

الگوریتم knn را در فایل ارائه شده در پیوست بصورت تابع knn پیاده سازی کردیم و برای سنجش فاصله از فاصله اقلیدسی و برای معیار انتخاب، ML را انتخاب پیاده سازی کردیم.

۶-۲. قسمت ۲

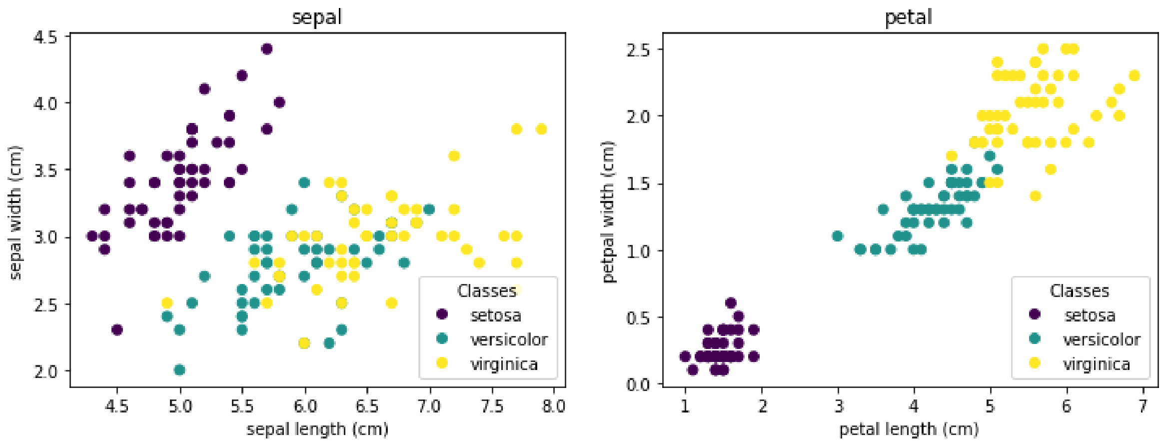
دیتاست iris را با تابع load_iris از کتابخانه sklearn فراخوانی کردیم. خصوصیات عددی این دیتاست اعم از بعد و تعداد سمپل ها و تعداد کلاس ها بصورت جدول زیر میباشد:

Classes	3
Samples per class	50
Samples total	150
Dimensionality	4
Features	real, positive

همانطور که از جدول بالا پیداست، این دیتاست شامل ۳ کلاس به نام های 'setosa', 'versicolor', 'virginica' میباشد. هر کلاس دارای ۴ ویژگی مثبت و حقیقی به نام های 'sepal length (cm)', 'sepal width (cm)', 'petal length (cm)', 'petal width (cm)' میباشد.

۶-۳. قسمت ۳

با توجه به ویژگی های ذکر شده، برای هر کدام از عرض و طول های petal و sepal نمودار scatter plot رسم میکنیم:



۶-۴. قسمت ۴

توسط دستور train_test_split کتابخانه sklearn، داده ها را به دو دسته آموزش و ارزیابی با نرخ ۲۰ درصد تقسیم میکنیم.

۶-۵. قسمت ۵

مدل پیاده سازی شده را با k=5 آموزش داده و سپس بر روی داده تست، ارزیابی میکنیم، دقت بدست آمده برابر است با ۹۳.۳۳ درصد.

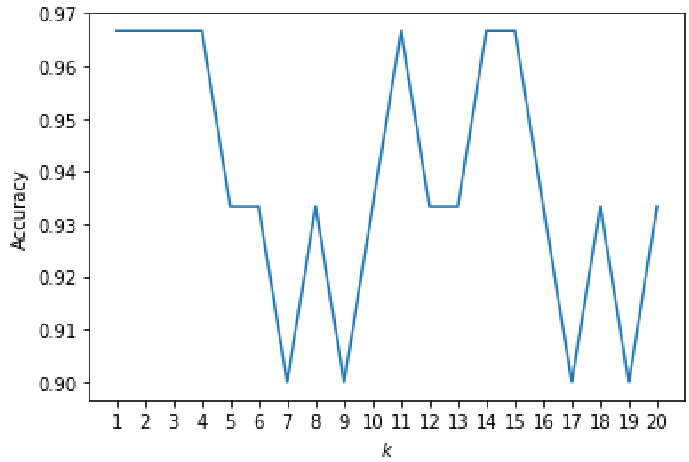
```
y_pred = knn(X_train, y_train, X_test, k=5, distance_fn=euclidean_distance, choice_fn=classification)
np.mean(y_pred - y_test == 0)*100
```

✓ 0.0s

```
93.33333333333333
```

۶-۶. قسمت ۶

در این قسمت برای اینکه بهترین انتخاب k را داشته باشیم، برای مجموعه k=1 تا k=20 نمودار accuracy را برای داده های آموزش و ارزیابی رسم کرده که بصورت زیر خواهد بود:

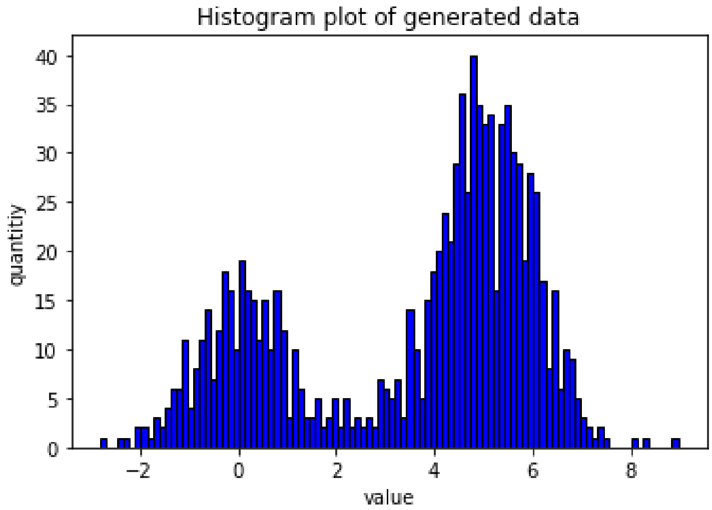


همانطور که مشاهده میشود، بیشتر دقت برای k های ۱ تا ۴ بدست می آید که برابر است با ۹۶.۶۶ ٪.

پاسخ ۷. شبیه سازی پنجره پارزن

۷-۱. قسمت ۱

در این قسمت با توجه به تکه کد داده شده، داده های مورد نظر تولید شده و سپس نمودار هیستوگرام این داده ها برای سنجش و مشاهده بهتر داده ها، رسم شده است.



۷-۲. قسمت ۲

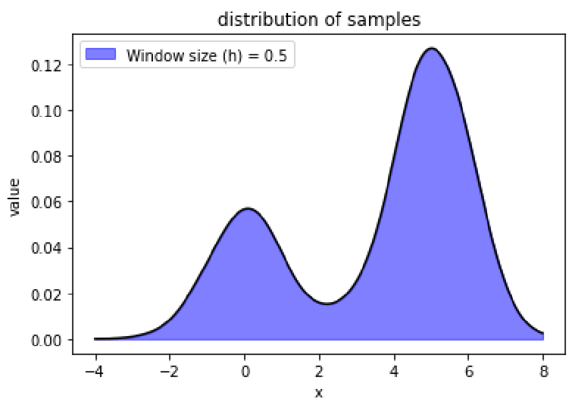
برای بدست آوردن توزیع داده ها، ابتدا یک کرنل گوسی به صورت زیر تعریف میکنیم:

```
Gaussian_kernel = lambda x : (1/(np.sqrt(2*np.pi)))*np.exp((x**2)/-2)
```

سپس با استفاده از فرمول زیر و مقدار h مورد نظر، توزیع را بدست می آوریم:

```
density = sum(Gaussian_kernel((xi - x_d)/h) for xi in X)/N
```

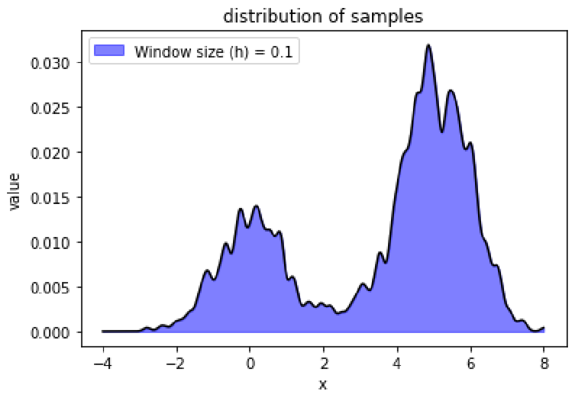
ما ابتدا طول پنجره را $h = 0.5$ قرار داده و توزیع را بصورت زیر رسم مینماییم:



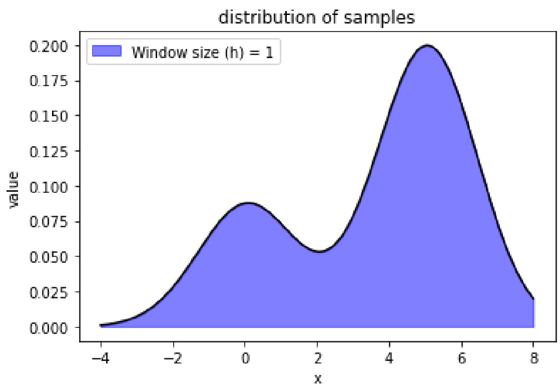
۷-۳. قسمت ۳

حال در این قسمت، اثرات مقادیر مختلف طول پنجره پارزن (h) را بررسی میکنیم و ۳ مقدار $h = 0.1,$ $h = 1, h = 10$ را مورد بررسی قرار میدهیم.

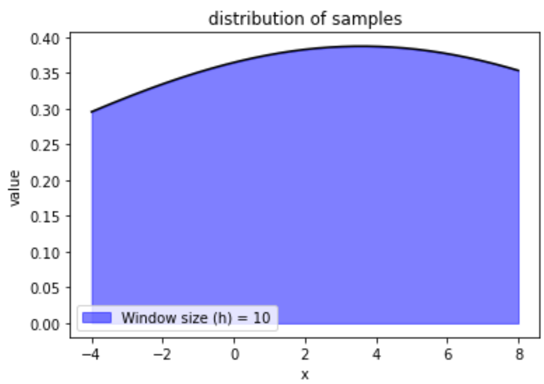
$h = 0.1$:



$h = 1$:



$h = 10$:

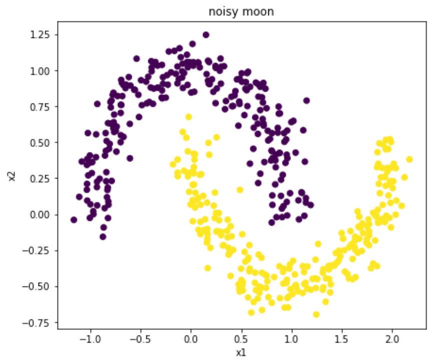


مشاهده میشود که با افزایش طول پنجره، توزیع نرم تر شده و تغییرات را نشان نمیدهد و از طرفی نیز در صورت کاهش طول، میزان نوسانات بسیار زیاد میشود و تغییرات شارپی را نشان میدهد.

پاسخ ۸. شبیه سازی GMM

۸-۱. قسمت ۱

توسط کد داده شده، داده ها را لود کرده و سپس آن ها را بصورت زیر رسم میکنیم:



۸-۲. قسمت ۲

حال برای هر کلاس بصورت جداگانه یک توزیع نرمال با روش بیزی تخمین میزنیم، در روش بیزی پارامتر های میانگین و کوواریانس را بدست آوردیم و مقادیر بدست آمده برای هر کلاس بصورت زیر میباشد:

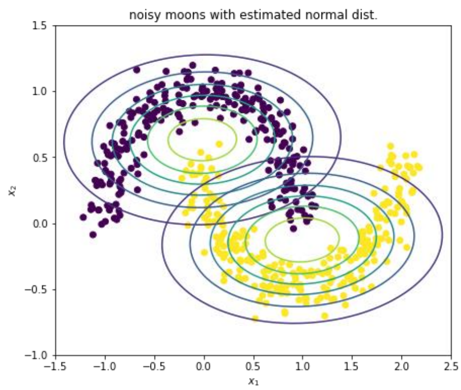
$$\hat{\mu}_1 = \begin{bmatrix} -0.01267522, & 0.63124533 \end{bmatrix}$$

$$\hat{\Sigma}_1 = \begin{bmatrix} 0.51361385 & 0.00710733 \\ 0.00710733 & 0.10861357 \end{bmatrix}$$

$$\hat{\mu}_2 = \begin{bmatrix} 0.99719469, & -0.12806477 \end{bmatrix}$$

$$\hat{\Sigma}_2 = \begin{bmatrix} 0.52159284 & 0.01385963 \\ 0.01385963 & 0.10351632 \end{bmatrix}$$

بر اساس پارامتر های بدست آمده بالا، توزیع های گوسی را بدست آورده و کانتور های آن را بصورت زیر رسم میکنیم:



۸-۳. قسمت ۳

همانطور که در قسمت قبل مشاهده شد، دو توزیع گوسی برای هرکدام از پارامتر ها کافی نبوده و خطای زیادی را در برمیگیرند. در این قسمت میخواهیم تعداد بیشتری توزیع گوسی را به هر کلاس فیت کنیم، بنابراین الگوریتم GMM را بر اساس روش EM پیاده سازی میکنیم.

در روش EM پارامتر ها در هر تکرار E و M بصورت زیر بدست می‌آیند:

تابع maximum likelihood بصورت زیر خواهد بود:

$$\ell(\phi, \mu, \Sigma) = \sum_{i=1}^{m} \log(p(x^{(i)}; \phi, \mu, \Sigma)) = \sum_{i=1}^{m} \log \sum_{z^{(i)}=1}^{k} p\left(x^{(i)} \mid z^{(i)}; \mu, \Sigma\right) p(z^{(i)}; \phi)$$

$$\ell(\phi, \mu, \Sigma) = \sum_{i=1}^{m} \log p\left(x^{(i)} \mid z^{(i)}; \mu, \Sigma\right) + \log p\left(z^{(i)}; \phi\right)$$

در نهایت با مشتق گیری پارامتر ها بصورت زیر بدست می‌آیند:

$$\phi_j = \frac{1}{m} \sum_{i=1}^{m} 1\{z^{(i)} = j\}$$

$$\mu_j = \frac{\sum_{i=1}^{m} 1\{z^{(i)} = j\} x^{(i)}}{\sum_{i=1}^{m} 1\left\{z^{(i)} = j\right\}}$$

$$\Sigma_j = \frac{\sum_{i=1}^{m} 1\{z^{(i)} = j\}(x^{(i)} - \mu_j)(x^{(i)} - \mu_j)^T}{\sum_{i=1}^{m} 1\{z^{(i)} = j\}}$$

برای GMM با توجه به فرمول های بالا می‌توان نوشت:

1. (E-step) For each i, j, set

$$w_j^{(i)} := p\left(z^{(i)} = j | x^{(i)}; \phi, \mu, \Sigma\right)$$

2. (M-step) Update the parameters

$$\phi_j := \frac{1}{m} \sum_{i=1}^{m} w_j^{(i)}$$

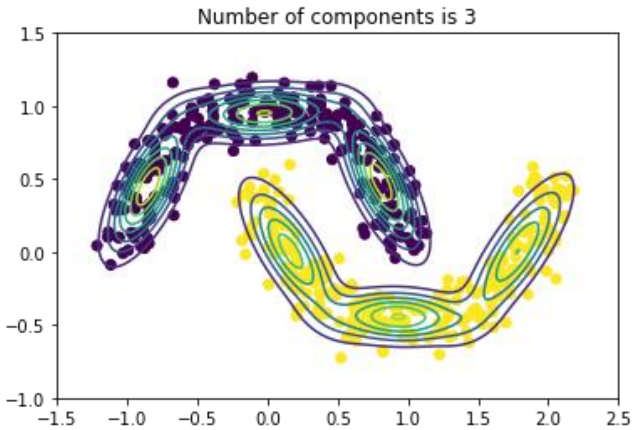
$$\mu_j := \frac{\sum_{i=1}^{m} w_j^{(i)} x^{(i)}}{\sum_{i=1}^{m} w_j^{(i)}}$$

$$\Sigma_j := \frac{\sum_{i=1}^{m} w_j^{(i)} \left(x^{(i)} - \mu_j\right) \left(x^{(i)} - \mu_j\right)^T}{\sum_{i=1}^{m} w_j^{(i)}}$$

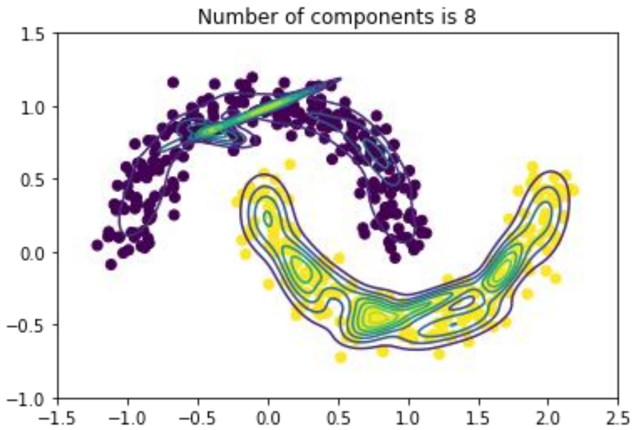
با توجه به فرمول های بالا، این الگوریتم را در کلاس GMM_EM پیاده کرده و در نهایت بر اساس تعداد مولفه های ۱ تا ۱۶، توزیع را بدست می آوریم.

کانتور های توزیع برای مولفه های ۳ و ۸ و ۱۶ بصورت زیر خواهد بود:

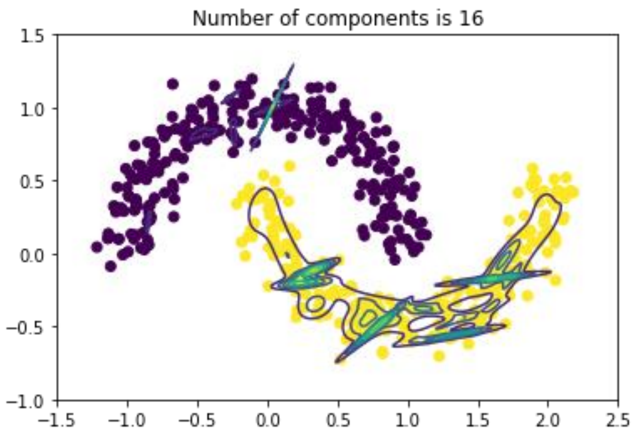
Number of components = 3:



Number of components = 8:



Number of components = 16:



۸-۴. قسمت ۴

حال برای اینکه برای هرکلاس تعداد مولفه های بهینه را بدست آوریم، از معیار های BIC و AIC استفاده کرده و برای هر مولفه این معیار ها را بدست می آوریم.

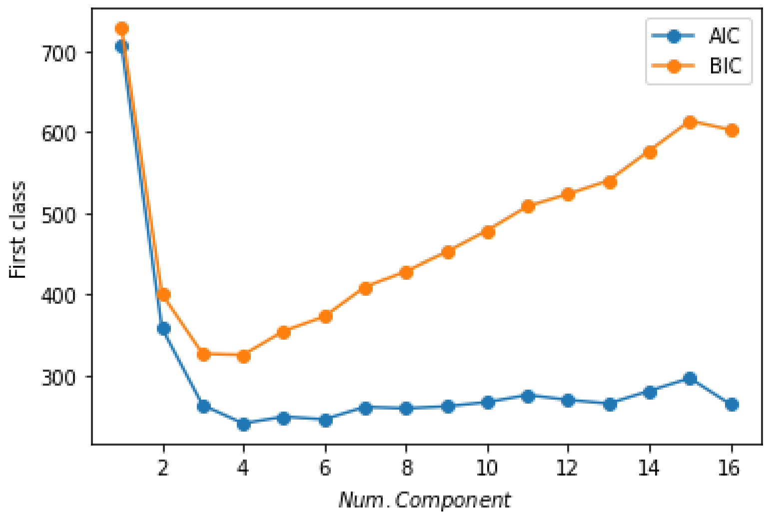
این معیار ها از توابع زیر بدست می آیند:

$$\begin{aligned} BIC = &\quad \log(m)\,p - 2\log\left(\widehat{L}\right) \\ AIC = &\quad 2p - 2\log\left(\widehat{L}\right) \end{aligned}$$

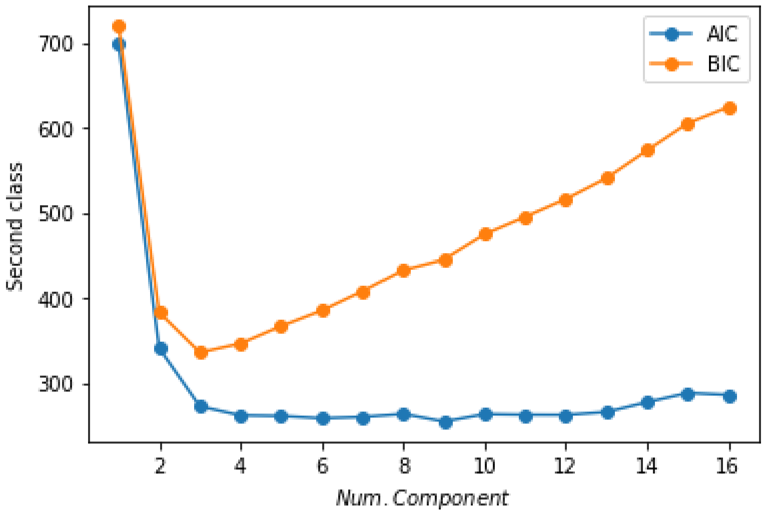
در روابط فوق m برابر است با تعداد نمونه ها، p برابر است با تعداد پارامترهای یادگرفته شده توسط مدل و $\hat{L}$ نیز بیشترین likelihood بدست آمده می‌باشد. توجه داریم که تعداد بهینه مولفه های نرمال براساس معیارهای AIC و BIC وقتی بدست می‌آید که این دو مقدارکمینه شده باشند.

این معیار ها نیز در کلاس نامبرده پیاده سازی شده و نمودار آن ها بر اساس تعداد مولفه ها از ۱ تا ۱۶ رسم شده تا بهترین تعداد مولفه بدست آید.

برای کلاس اول داریم:



برای کلاس دوم داریم:



با توجه به معیار گفته شده برای انتخاب بهینه تعداد مولفه های گاوسی و نمودارهای فوق، بهترین تعداد مولفه برای هرکلاس ۴ می‌باشد.

برای ۴ مولفه، کانتور توزیع برای هرکلاس بصورت زیر میباشد: