به نام خدا

دانشگاه تهران

پردیس دانشکده‌های فنی

دانشکده برق و کامپیوتر

شناسائی الگو

تمرین شماره ۲

نام و نام خانوادگی

سجاد پاکدامن ساوجی

شماره دانشجویی

۸۱۰۱۹۵۵۱۷

فهرست

چکیده

در این تمرین با مباحثی از درس مانند MAP ، ML و EM آشنا شدیم .همچنان به صورت امتیازی با hidden markov models آشنا شدیم.

سوال 1

Pattern Recognition Assignment 2

Problem 1)

I. $$f(x^q|\mu) = \frac{1}{(2\pi)^{\frac{n}{2}}\,|\underline{\underline{\Sigma}}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x^q-\mu)^T \underline{\underline{\Sigma}}_{\circ}^{-1}(x^q-\mu)\right\}$$

$$f(\mu) = \frac{1}{(2\pi)^{\frac{n}{2}}\,|\underline{\underline{\Sigma}}_{\circ}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mu-m_\circ)^T \underline{\underline{\Sigma}}_{\circ}^{-1}(\mu-m_\circ)\right\}$$

$$\hat{\theta}_{map} = \underset{\theta}{\operatorname{argmax}}\; f(\theta|\mathcal{X}) = \underset{\theta}{\operatorname{argmax}}\; f(\mathcal{X}|\theta)\, f(\theta)$$

$$= \underset{\theta}{\operatorname{argmax}}\left(\prod_{q=1}^{Q} f(x^q|\theta)\right) f(\theta) \xrightarrow{\ln} \underset{\theta}{\operatorname{argmax}} \sum_{q=1}^{Q} \ln f(x^q|\theta) + \ln f(\theta)$$

$$L(\theta) = \sum_{q=1}^{Q}\left[-\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\underline{\underline{\Sigma}}| - \frac{1}{2}(x^q-\mu)^T \underline{\underline{\Sigma}}^{-1}(x^q-\mu)\right] - \frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\underline{\underline{\Sigma}}_\circ| - \frac{1}{2}(\mu-m_\circ)^T \underline{\underline{\Sigma}}_\circ^{-1}(\mu-m_\circ)$$

$$\frac{\partial L(\theta)}{\partial \theta} = 0 \Rightarrow \sum_{q=1}^{Q}\left[-\frac{1}{2}\left(2\underline{\underline{\Sigma}}^{-1}\mu - 2\underline{\underline{\Sigma}}^{-1}x^q\right)\right] - \frac{1}{2}\left(2\underline{\underline{\Sigma}}_\circ^{-1}\mu - 2\underline{\underline{\Sigma}}_\circ^{-1}m_\circ\right) = 0$$

$$\Rightarrow \left(Q\,\underline{\underline{\Sigma}}^{-1} + \underline{\underline{\Sigma}}_\circ^{-1}\right)\mu = \underline{\underline{\Sigma}}^{-1}\sum_{q=1}^{Q} x^q + \underline{\underline{\Sigma}}_\circ^{-1} m_\circ$$

$$\Rightarrow \hat{\mu}_{map} = \left(Q\,\underline{\underline{\Sigma}}^{-1} + \underline{\underline{\Sigma}}_\circ^{-1}\right)^{-1}\left[\underline{\underline{\Sigma}}^{-1}\sum_{q=1}^{Q} x^q + \underline{\underline{\Sigma}}_\circ^{-1} m_\circ\right]$$

بدیهی است که نحوه تغییر پایه (نگاشت خطی) با پاسخ مسئله رابطه مستقیم دارد.
اگر نگاشت خطی یک به یک، پوشا نباشد روابط بین μ، $\tilde{\mu}$ برقرار نخواهد بود

$$\tilde{\mu} = E\{\tilde{x}\} = E\{Ax\} = A\,E\{x\} = A\mu \Rightarrow \tilde{m}_\circ = A m_\circ$$

$$\tilde{\underline{\underline{\Sigma}}} = E\left\{(\tilde{x}-\tilde{\mu})(\tilde{x}-\tilde{\mu})^T\right\} = E\left\{A(x-\mu)(x-\mu)^T A^T\right\} = A\,\underline{\underline{\Sigma}}\,A^T$$

$$\Rightarrow \tilde{\underline{\underline{\Sigma}}}_\circ = A\,\underline{\underline{\Sigma}}_\circ\,A^T$$

بدون از دست رفتن کلیات فرض می‌کنیم که A $\leftarrow$ unitary

$$\tilde{\mu} = \left(Q\,\tilde{\Sigma}^{-1} + \tilde{\Sigma}_{\cdot}^{-1}\right)^{-1}\left[\tilde{\Sigma}^{-1}\sum_{q=1}^{Q}\tilde{x}^{q} + \tilde{\Sigma}_{\cdot}^{-1}\,\tilde{m}_{\cdot}\right]$$

$$Q\,\tilde{\Sigma}^{-1} + \tilde{\Sigma}_{\cdot}^{-1} = Q\,A\,\Sigma^{-1}A^{T} + A\,\Sigma_{\cdot}\,A^{T} = A\left[Q\,\Sigma^{-1} + \Sigma_{\cdot}^{-1}\right]A^{T}$$

$$\Rightarrow \left(Q\,\tilde{\Sigma}^{-1} + \tilde{\Sigma}_{\cdot}^{-1}\right)^{-1} = \left(A\left[Q\,\Sigma^{-1} + \Sigma_{\cdot}^{-1}\right]A^{T}\right)^{-1} = A\left[Q\,\Sigma^{-1} + \Sigma_{\cdot}^{-1}\right]^{-1}A^{T}$$

$$\tilde{\Sigma}^{-1}\sum_{q=1}^{Q}\tilde{x}^{q} + \tilde{\Sigma}_{\cdot}^{-1}\,\tilde{m}_{\cdot} = A\,\Sigma^{-1}\underbrace{A^{T}\left(A}_{I}\sum_{q=1}^{Q}x^{q}\right) + A\,\Sigma_{\cdot}^{-1}\underbrace{A^{T}A}_{I}m_{\cdot}$$

$$= A\left(\Sigma^{-1}\sum_{q=1}^{Q}x^{q} + \Sigma_{\cdot}^{-1}m_{\cdot}\right)$$

$$\Rightarrow \tilde{\mu} = A\left[Q\,\Sigma^{-1} + \Sigma_{\cdot}^{-1}\right]^{-1}\underbrace{A^{T}A}_{I}\left[\Sigma^{-1}\sum_{q=1}^{Q}x^{q} + \Sigma_{\cdot}^{-1}m_{\cdot}\right] = A\mu$$

پس درحالتی که ماتریس A anitary باشد روابط بدست آمده برای μ ، در مورد $\tilde{\mu}$ هم صادق هستند.

II. $f(x^{q}|\mu) = \dfrac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{\dfrac{-(x-\mu)^{2}}{2\sigma^{2}}\right\}$

$f(\mu) = \mu\,\exp\left\{\dfrac{-\mu^{2}}{2\sigma_{\mu}^{2}}\right\}$

$$\hat{\theta}_{MAP} = \arg\max_{\theta}\underbrace{\sum_{q=1}^{Q}\ln f(x^{q}|\mu) + \ln f(\mu)}_{L(\theta)}$$

$$L(\theta) = \sum_{q=1}^{Q}\left(-\frac{1}{2}\ln 2\pi - \ln\sigma - \frac{(\mu-x^{q})^{2}}{2\sigma^{2}}\right) - \ln\sigma_{\mu}^{2} + \ln\mu - \frac{\mu^{2}}{2\sigma_{\mu}^{2}}$$

$$\frac{\partial L(\theta)}{\partial\theta} = 0 \Rightarrow \sum_{q=1}^{Q}\frac{2(\mu-x^{q})}{2\sigma^{2}} - \frac{1}{\mu} + \frac{2\mu}{2\sigma_{\mu}^{2}} = 0$$

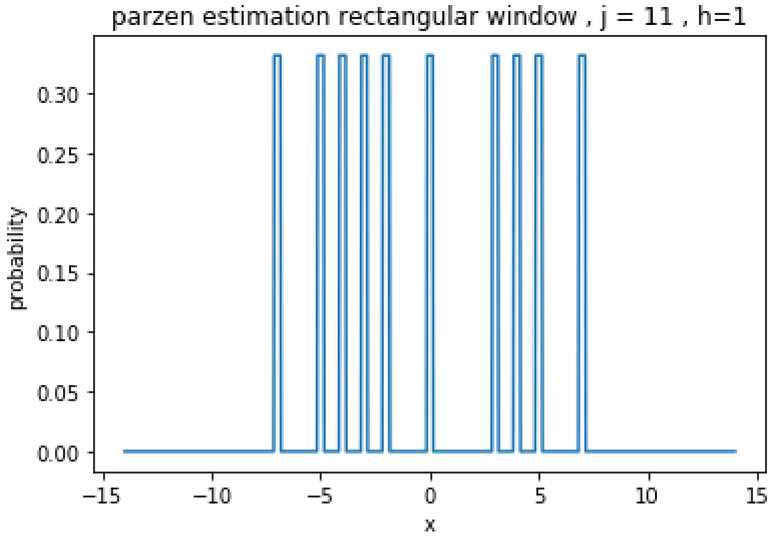
$$\Rightarrow \underbrace{-\sum_{q=1}^{Q}\frac{x^{q}}{\sigma^{2}}}_{Z} - \frac{1}{u} + \underbrace{\left(\frac{Q}{\sigma^{2}} + \frac{1}{\sigma_{\mu}^{2}}\right)}_{R}\mu = 0 \quad \Rightarrow \mu^{2} - \frac{Z}{R}\mu - \frac{1}{R} = 0$$

$$\Rightarrow \mu = \frac{\frac{Z}{R} \pm \sqrt{\left(\frac{Z}{R}\right)^{2} + \frac{4}{R}}}{2} = \frac{Z}{2R}\left(1 \pm \sqrt{1 + \frac{4R}{Z^{2}}}\right)$$

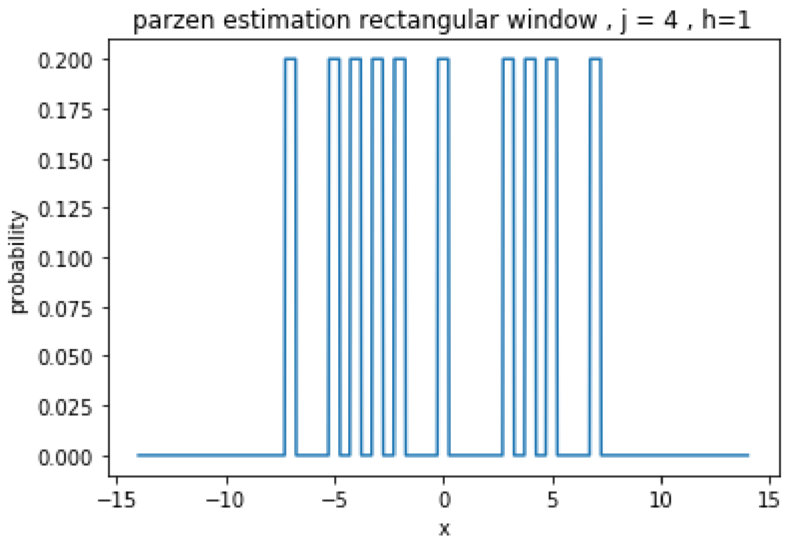
سوال 2

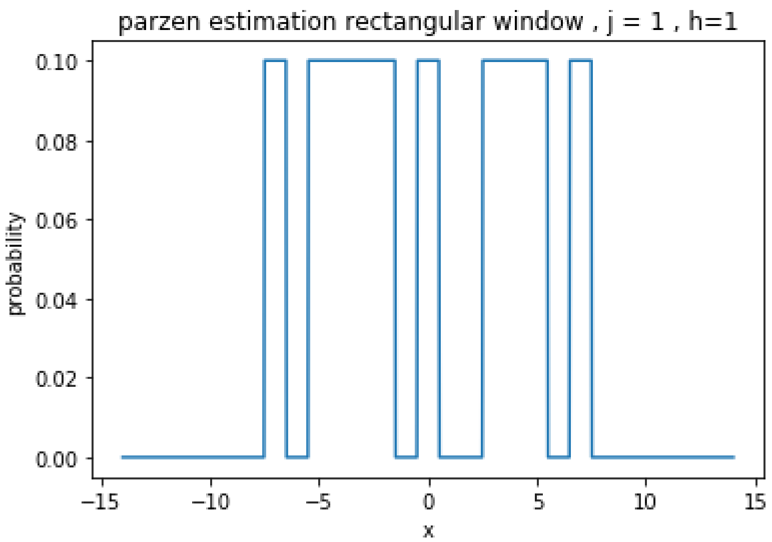
محاسبات مربوط به این سوال در فایل p2.inpy قرار دارد و پلات ها و نتایج آن در زیر آمده است.

قسمت ۱ (a)

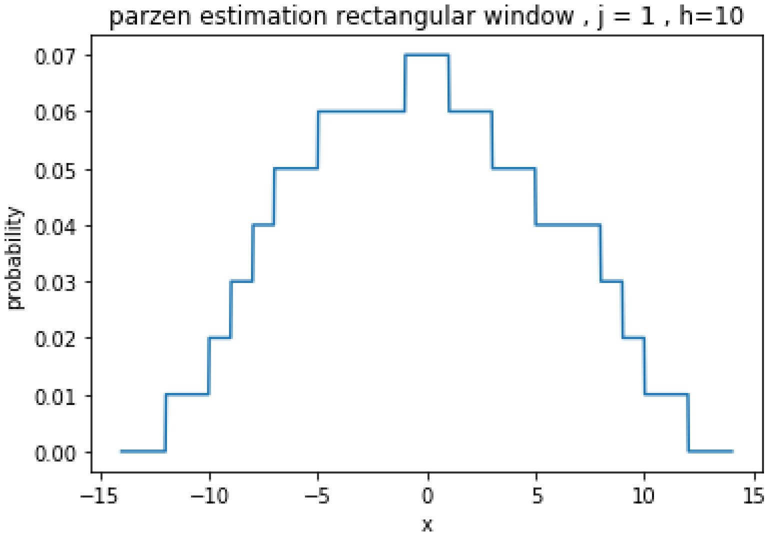

parzen estimation rectangular window , j = 11 , h=1
probability
x

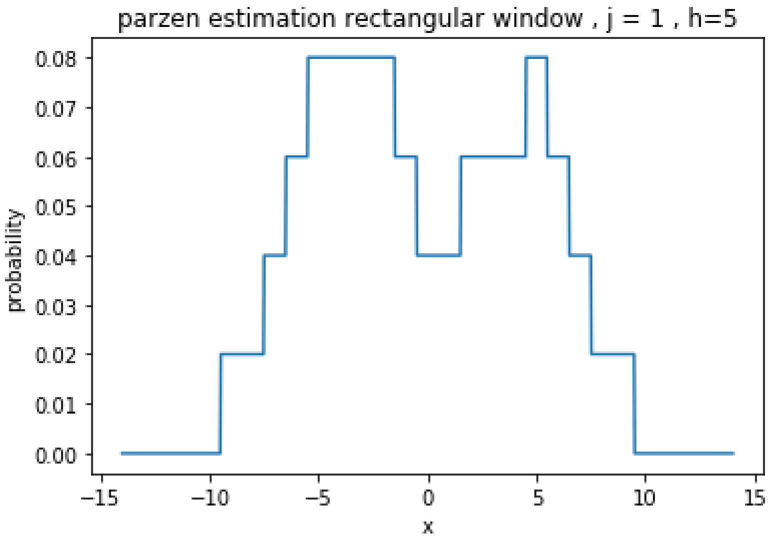
a)


parzen estimation rectangular window , j = 4 , h=1
probability
x

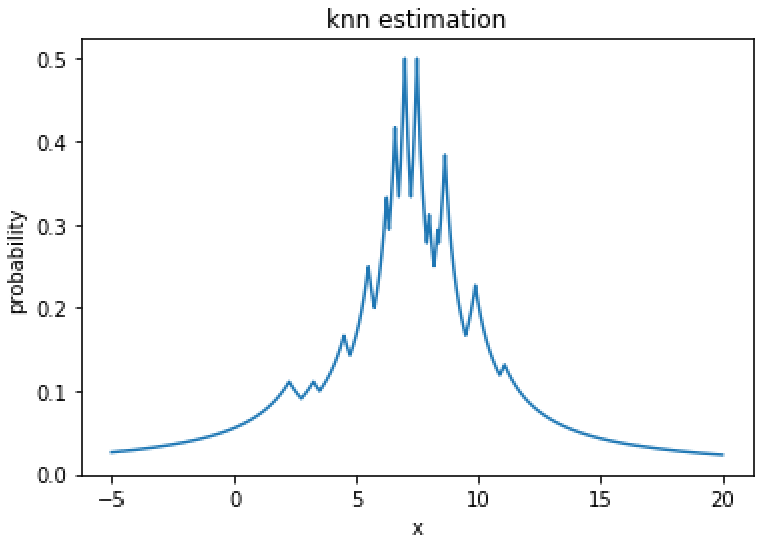

parzen estimation rectangular window , j = 1 , h=1
probability
x

قسمت ۱ (b)


parzen estimation rectangular window , j = 1 , h=10
probability
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
−15
−10
−5
0
5
10
15
x


parzen estimation rectangular window , j = 1 , h=5
probability
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
−15
−10
−5
0
5
10
15
x

قسمت ۲) بخش ۱)



knn estimate (2 , k = 4) = 0.1

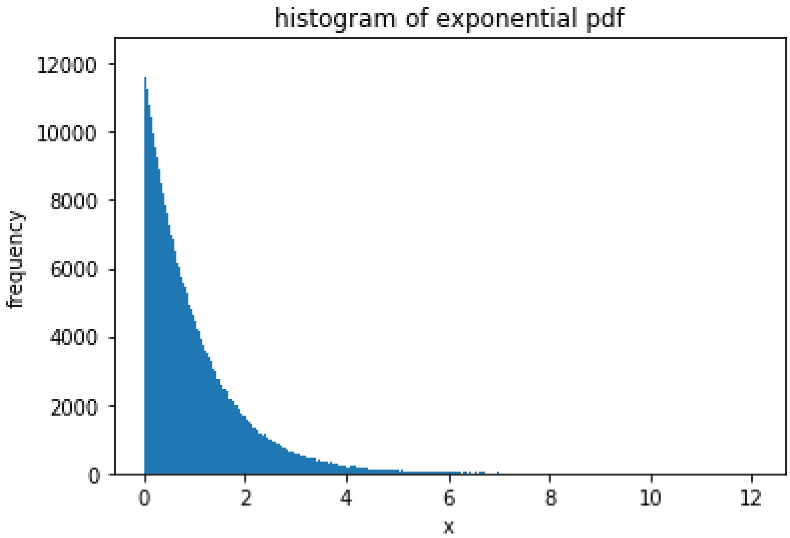
knn estimate (4 , k = 4) = 0.125

knn estimate (6 , k = 4) = 0.25

knn estimate (8 , k = 4) = 0.312

knn estimate (10 , k = 4) = 0.208

قسمت ۲) بخش۲) a)



قسمت ۲) بخش ۲) b)

std_10 = 1.70

std_100 = 0.9602

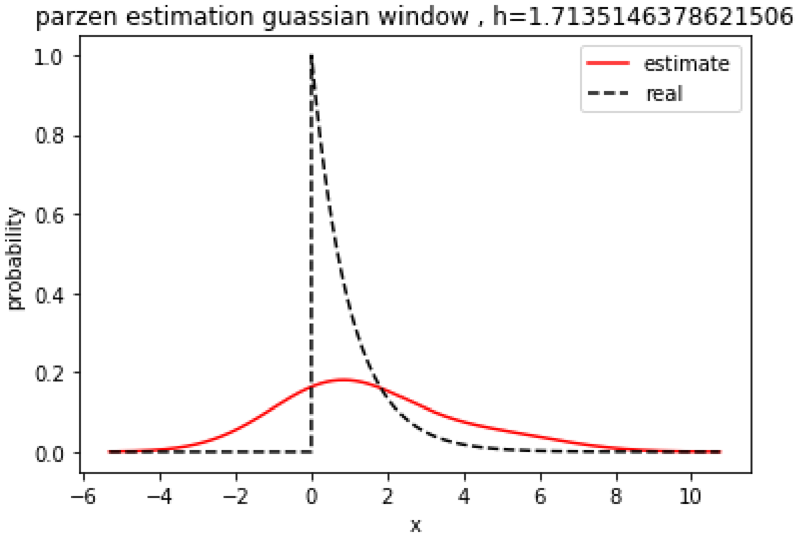
std_1000 = 0.9665

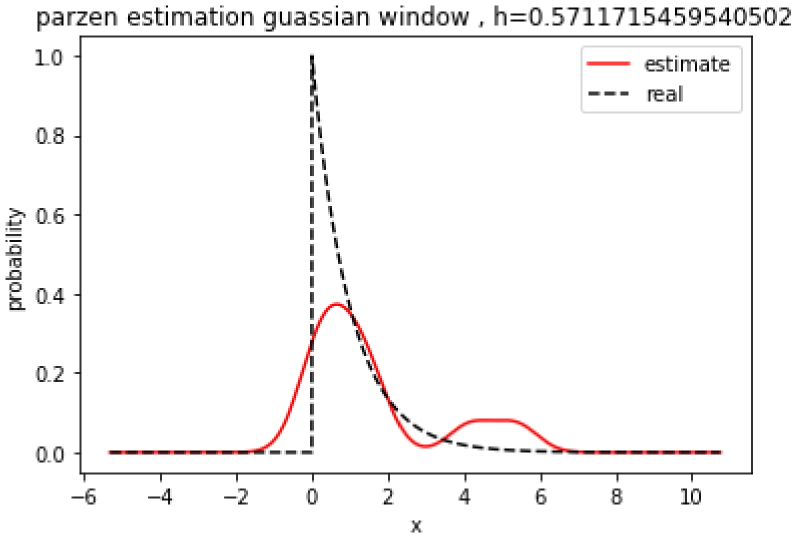
h*_10 = 0.5711

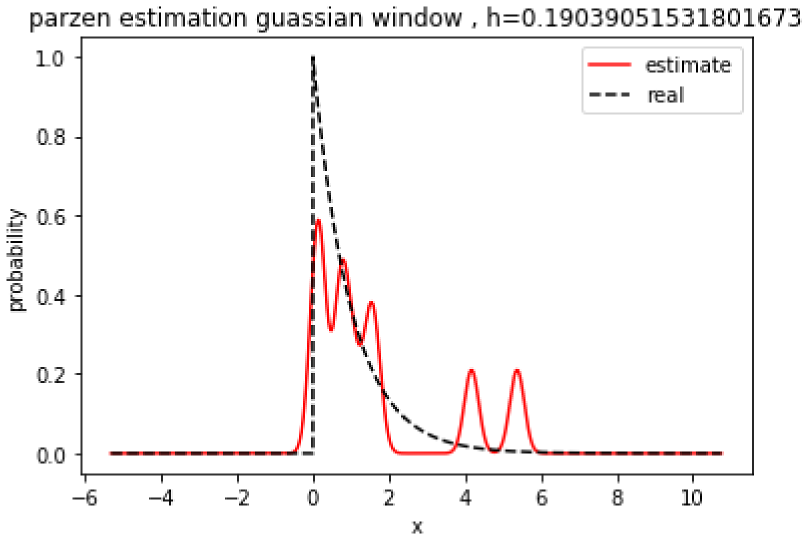
h*_100 = 0.1017

h*_1000 = 0.032

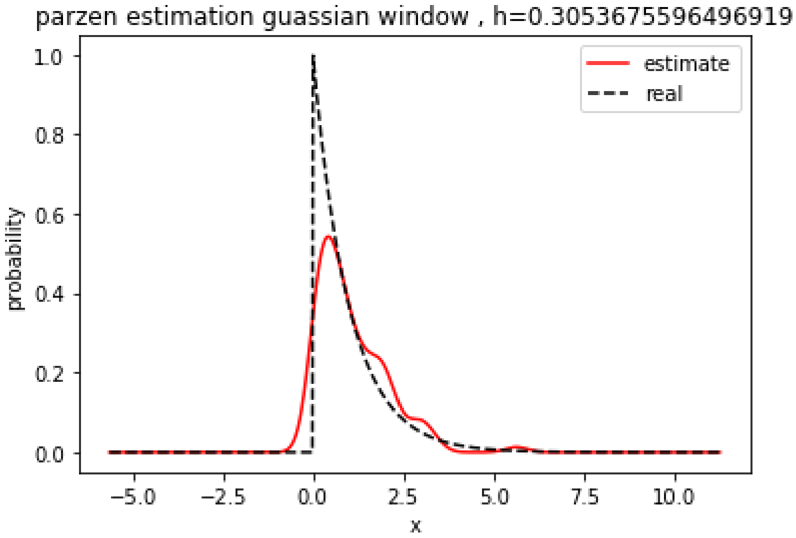
برای تعداد داده ۱۰ داریم :

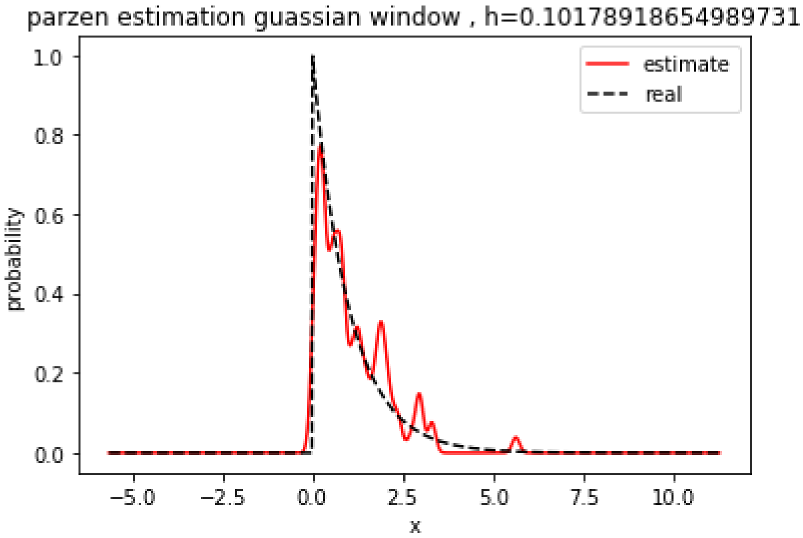

parzen estimation guassian window , h=1.7135146378621506
estimate
real
probability
x

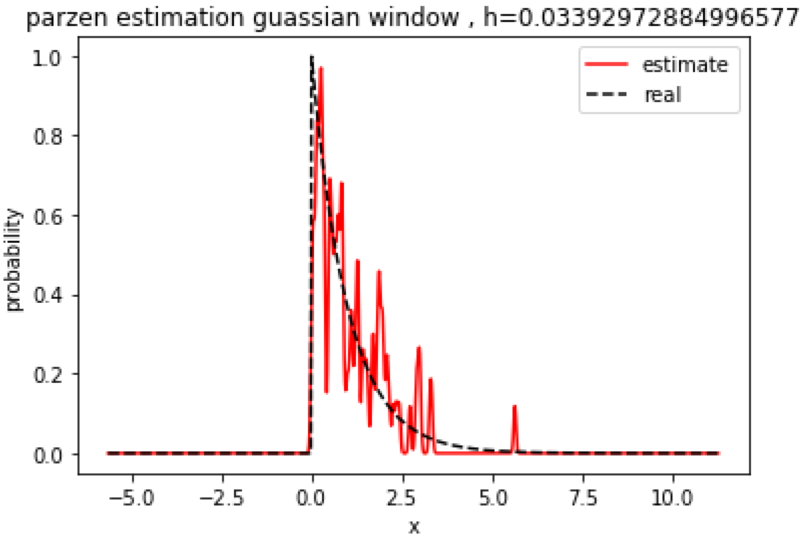

parzen estimation guassian window , h=0.5711715459540502
estimate
real
probability
x


parzen estimation guassian window , h=0.19039051531801673
estimate
real
probability
x

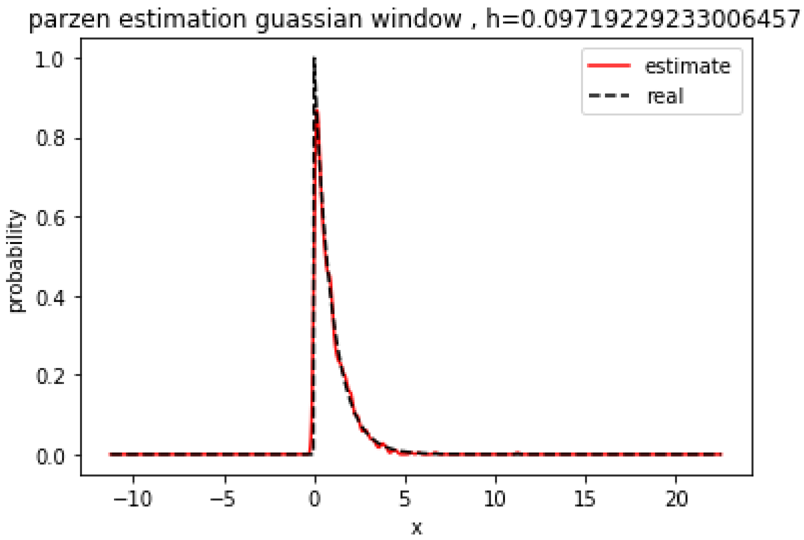
برای تعداد داده ۱۰۰ داریم:

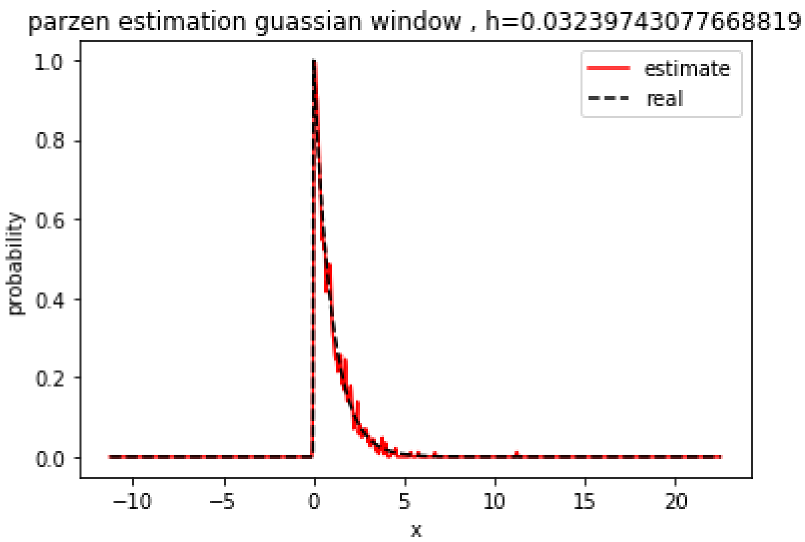


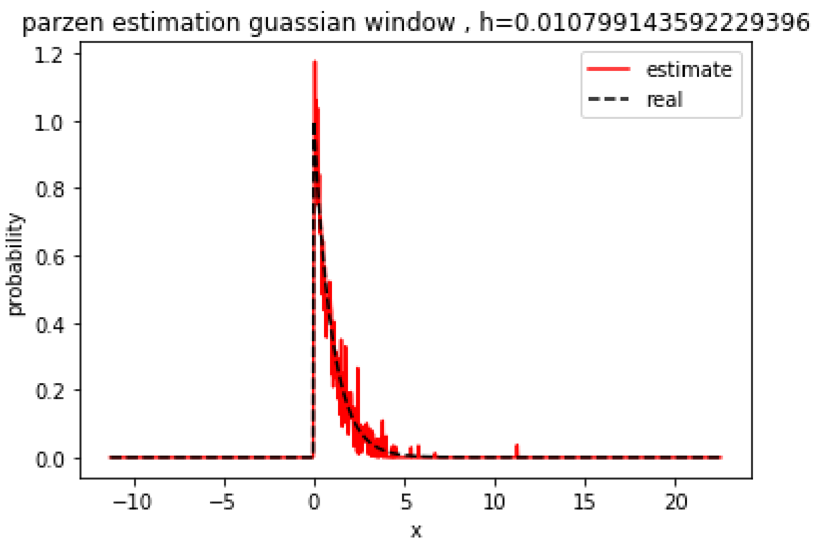




برای تعداد داده ۱۰۰۰ داریم:







نتیجه گیری :خود kernel تاثیر مستقیم روی کیفیت تخمین دارد و همان طور که مشاهده کردیم kernel گوسی از kernel مربعی بهتر عمل کرد زیرا این کرنل وزن یکسانی به داده ها نمیدهد.

نقش N به این صورت است که هر چه N بیشتر باشد ما تعداد داده های بیشتری داریم و در نتیجه کیفیت تخمین ما بهتر خواهد شد. و در حالت حدی که n خیلی بزرگ است در این صورت دیگر مهم نیست که از چه کرنلی استفاده کنیم در هر حال کیفیت تخمین مناسب است.

نقش h در حقیقت یک اسکیل ساده است. از آن جایی که در حوزه فرکانس کرنل و توزیع با یک دیگر کانوالو میشوند ، پس تفسیر دیگری از h همان پهنای باند در حوزه فرکانس است. h در این مفهوم اجازه نمیدهد تغییرات زیاد وارد تخمین شود و در نتیجه تابع تخمین زده شده نرم تر خواهد شد.

سوال 3

Problem 3) ML-Estimation

I.

a. $\hat{\theta}_{ML} = \underset{\theta}{\operatorname{argmax}}\ f(\chi|\theta) \qquad \hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}}\ f(\chi|\theta)\ f(\theta)$

باتوجه به روابط بالا نقش $f(\theta)$ regulization در optimization است. درحالت کلی زمانی این regulization
تاثیری ندارد که $f(\theta)$ اطلاعاتی به مسئله اضافه نکند

برای مثال برای چگالی گوسی با تخمین $\hat{\theta}_{MAP}$ به صورت زیر است.

$$\hat{\mu}_{MAP} = \frac{\left(\frac{\sigma_\mu}{\sigma}\right)^2 \sum_{q=1}^{Q} x^q + \mu_0}{Q\left(\frac{\sigma_\mu}{\sigma}\right)^2 + 1}$$

$$\frac{\sigma_\mu}{\sigma} \gg 1 \longrightarrow \hat{\mu}_{MAP} = \hat{\mu}_{ML}$$

$$\frac{\sigma_\mu}{\sigma} \ll 1 \longrightarrow \hat{\mu}_{MAP} = \mu_0$$

b. $\hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}}\ f(\chi|\theta) f(\theta) \xrightarrow{\ln} \hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}}\ \ln^{f(\chi|\theta)} + \ln^{f(\theta)}$

$\xrightarrow{\text{iid}} \underset{\theta}{\operatorname{argmax}}\ \ln^{\prod_{q=1}^{Q} f(x^q|\theta)} + \ln^{f(\theta)} \xrightarrow{\text{naive}} \underset{\theta}{\operatorname{argmax}}\ \ln^{\prod_{q=1}^{Q}\prod_{k=1}^{N} f(x_k^q|\theta)} + \ln^{f(\theta)}$

اگر ابعاد بردار ویژگی conditional independance دارند

$$\Rightarrow \hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}} \sum_{q=1}^{Q}\sum_{k=1}^{N} f(x_k^q|\theta) + \ln^{f(\theta)}$$

با فرض naive bays تخمین MAP به صورت بالا تبدیل خواهد شد.

II.

a) $f(x_k;\theta) = \frac{1}{\sigma x_k\sqrt{2\pi}} \exp\left\{-\frac{(\ln^{x_k}-\theta)^2}{2\sigma^2}\right\} \qquad x_k > 0$

$$\Rightarrow \hat{\theta}_{ML} = \underset{\theta}{\operatorname{argmax}} \sum_{q=1}^{Q} \ln\left(\frac{1}{\sigma x^q\sqrt{2\pi}}\right) - \frac{(\ln^{x^q}-\theta)^2}{2\sigma^2} = \sum_{q=1}^{Q} -\ln^{\sigma} - \ln^{x^q} - \frac{1}{2}\ln^{2\pi} - \frac{(\ln^{x^q}-\theta)^2}{2\sigma^2}$$

$$\frac{\partial l(\theta)}{\partial\theta} = \sum_{q=1}^{Q} \frac{-2(\theta - \ln^{x^q})}{2\sigma^2} = 0 \Rightarrow \sum_{q=1}^{Q} \frac{-\theta}{\sigma^2} + \frac{\ln^{x^q}}{\sigma^2} = 0$$

$$\Rightarrow Q\theta = \sum_{q=1}^{Q} \ln^{x^q} \Rightarrow \hat{\theta}_{ML} = \frac{1}{Q}\sum_{q=1}^{Q} \ln^{x^q}$$

b. $f(x^q|\theta) = \frac{1}{\Gamma(\alpha)} \theta^{\alpha} x^{q^{\alpha-1}} \exp\{-\theta x^q\}$

$$\hat{\theta}_{ML} = \arg\max_{\theta} \sum_{q=1}^{Q} -\ln^{\Gamma(\alpha)} + \alpha \ln^{\theta} + (\alpha-1) \ln^{x^q} - \theta x^q$$

$$\frac{\partial L(\theta)}{\partial \theta} = \sum_{q=1}^{Q} \alpha \times \frac{1}{\theta} - x^q = 0 \Rightarrow \frac{\alpha Q}{\theta} = \sum_{q=1}^{Q} x^q \Rightarrow \hat{\theta}_{ML} = \frac{\alpha Q}{\sum_{q=1}^{Q} x^q}$$

c. $f(x^q|\theta) = \sqrt{\theta}\, x^{q^{\sqrt{\theta}-1}}$

$$\hat{\theta}_{ML} = \arg\max_{\theta} \sum_{q=1}^{Q} \frac{1}{2} \ln^{\theta} + (\sqrt{\theta}-1) \ln^{x^q}$$

$$\frac{\partial L(\theta)}{\partial \theta} = \sum_{q=1}^{Q} \frac{1}{2\theta} + \frac{+\ln^{x^q}}{2\sqrt{\theta}} = 0 \overset{\theta > 0}{\Rightarrow} \sum_{q=1}^{Q} \frac{1 + \sqrt{\theta} \ln^{x^q}}{2\theta} = 0$$

$$\Rightarrow Q + \sqrt{\theta} \sum_{q=1}^{Q} \ln^{x^q} = 0 \Rightarrow \sqrt{\theta} = \frac{-Q}{\sum_{q=1}^{Q} \ln^{x^q}} \Rightarrow \hat{\theta}_{ML} = \left(\frac{-Q}{\sum_{q=1}^{Q} \ln^{x^q}} \right)^2$$

d. $f(x^q|\theta) = \theta^2 x^q \exp\{-\theta x^q\} u(x^q) \qquad \theta \geq 0, \; x^q \geq 0$

$$\hat{\theta}_{ML} = \arg\max_{\theta} \sum_{q=1}^{Q} 2\ln^{\theta} + \ln^{x^q} - \theta x^q + \ln^{u(x^q)}$$

$$\frac{\partial L(\theta)}{\partial \theta} = \sum_{q=1}^{Q} \frac{2}{\theta} - x^q = 0 \Longrightarrow \hat{\theta}_{ML} = \frac{2Q}{\sum_{q=1}^{Q} x^q}$$

سوال 4

problem4) Expectation Maximization

$$f(x^q|\theta,j)=\frac{1}{\sqrt{2\pi}\sigma_j}\exp\left\{-\frac{(x^q-\mu_j)^2}{2\sigma_j^2}\right\}$$

$$Q(\theta|\hat{\theta}(t),\mathcal{X})=\sum_{q=1}^{Q}\sum_{j=1}^{J}p(j|\hat{\theta}(t),x^q)\left[\frac{-(x^q-\mu_j)^2}{2\sigma_j^2}-\frac{n}{2}\ln^{2\pi\sigma_j^2}+\ln P_j\right]$$

a). $\sum_{j=1}^{J}P_j=1$

$$\Rightarrow L(\theta,\lambda)=\sum_{q=1}^{Q}\sum_{j=1}^{J}P(j|\hat{\theta}(t),x^q)\left[\frac{-(x^q-\mu_j)^2}{2\sigma_j^2}-\frac{n}{2}\ln^{2\pi\sigma_j^2}+\ln^{P_j}\right]-\lambda\left[\sum_{i=1}^{J}P_i-1\right]$$

$$\frac{\partial L(\theta,\lambda)}{\partial\mu_j}=\sum_{q=1}^{Q}P(j|\hat{\theta}(t),x^q)\left[\frac{-(\mu_j-x^q)}{\sigma_j^2}\right]=0\Rightarrow\sum_{q=1}^{Q}P(j|\hat{\theta}(t),x^q)\left[\mu_j-x^q\right]=0$$

$$\Rightarrow\mu_j\sum_{q=1}^{Q}P(j|x^q,\hat{\theta}(t))=\sum_{q=1}^{Q}P(j|x^q,\hat{\theta}(t))\,x^q\Rightarrow\hat{\mu}_{j_{(t+1)}}=\frac{\sum_{q=1}^{Q}P(j|x^q,\hat{\theta}(t))\,x^q}{\sum_{q=1}^{Q}P(j|x^q,\hat{\theta}(t))}$$

$$\frac{\partial L(\theta,\lambda)}{\partial\sigma_j^2}=\sum_{q=1}^{Q}P(j|\hat{\theta}(t),x^q)\left[\frac{+(x^q-\mu_j)^2}{2(\sigma_j^2)^2}-\frac{n}{2}\times\frac{1}{2\pi\sigma_j^2}\times2\sigma_j\times2\pi\right]=0$$

$$\sum_{q=1}^{Q}\frac{P(j|\hat{\theta}(t),x^q)}{(\sigma_j^2)^2}\left[\frac{(x^q-\mu_j)^2}{2}-\frac{n}{4\pi}\times2\sigma_j\times\sigma_j^2\times2\pi\right]=0$$

$$\sum_{q=1}^{Q}\frac{P(j|\hat{\theta}(t),x^q)}{(\sigma_j^2)^2}\left[\frac{(x^q-\mu_j)^2}{2}-\frac{n}{2\pi}\sigma_j^3\times2\pi\right]=0$$

$$\Rightarrow\sum_{q=1}^{Q}P(j|\hat{\theta}(t),x^q)\left[\frac{(x^q-\mu_j)^2}{\sigma_j^3}-n\frac{\sigma_j^2}{\sigma_j^3}\right]=0\Rightarrow\sum_{q=1}^{Q}P(j|\hat{\theta}(t),x^q)\left[(x^q-\mu_j)^2-n\sigma_j^2\right]=0$$

$$\Rightarrow nQ\sigma_j^2\times\sum_{q=1}^{Q}P(j|\hat{\theta}(t),x^q)=\sum_{q=1}^{Q}P(j|\hat{\theta}(t),x^q)(x^q-\mu_j)^2$$

$$\Rightarrow\hat{\sigma}_{j_{(t+1)}}^2=\frac{\sum_{q=1}^{Q}P(j|\hat{\theta}(t),x^q)(x^q-\mu_j)^2}{nQ\sum_{q=1}^{Q}P(j|\hat{\theta}(t),x^q)}$$

$$\frac{\partial L(\theta,\lambda)}{\partial P_j}=\left[\sum_{q=1}^{Q}P(j|\hat{\theta}(t),x^q)\times\frac{1}{P_j}\right]-\lambda=0\Rightarrow\frac{1}{P_j}\sum_{q=1}^{Q}P(j|\hat{\theta}(t),x^q)=\lambda\Rightarrow P_j=\frac{\sum_{q=1}^{Q}P(j|\hat{\theta}(t),x^q)}{\lambda}$$

$$\frac{\partial L(\theta,\lambda)}{\partial L}=0\Rightarrow\sum_{i=1}^{J}P_j=1\Rightarrow\lambda=\sum_{j=1}^{J}\sum_{q=1}^{Q}P(j|x^q,\hat{\theta}(t))=Q\Rightarrow P_{j_{(t+1)}}=\frac{\sum_{q=1}^{Q}P(j|\hat{\theta}(t),x^q)}{Q}$$

b) $p(j|\hat{\theta}(t), x^q) = ? \xrightarrow{\text{bays}} \dfrac{p(x^q|\hat{\theta}(t), j)\, p(j|\hat{\theta}(t))}{p(x^q|\hat{\theta}(t))} = \dfrac{p(x^q|\hat{\theta}(t), j)\, p(j|\hat{\theta}(t))}{\sum_{i=1}^{J} p(x^q|\hat{\theta}(t), i)\, p(i|\hat{\theta}(t))}$

c) $\theta = [\mu_j, \sigma_j, p_j, \ldots\ldots,]^T \Rightarrow \theta = [\mu_1, \sigma_1, p_1, \ldots\ldots, \mu_J, \sigma_J, p_J]^T$

d) $y = \begin{pmatrix} \underline{x} \\ \underline{x}_m \end{pmatrix}$ $\quad \underline{x}_m \to x$ missing $\qquad \mathcal{X} = \{(y^q)\}_{q=1}^{Q} \to$ unsupervised $\qquad f(y|\theta) = f_x(\underline{x}|\theta)\, f(\underline{x}_m)$

$$\hat{\theta}_{ML} = \underset{\theta}{\arg\max}\ f_y(y|\theta) \xrightarrow{\text{iid}} \hat{\theta}_{ML} = \underset{\theta}{\arg\max} \prod_{q=1}^{Q} f_y(y^q|\theta) = \underset{\theta}{\arg\max} \prod_{q=1}^{Q} f(\underline{x}^q|\theta)\, f(\underline{x}_m^q)$$

$$\xrightarrow{\ln} \hat{\theta}_{ML} = \underset{\theta}{\arg\max} \sum_{q=1}^{Q} \ln^{f(\underline{x}^q|\theta)} + \ln^{f(\underline{x}_m^q)} \longrightarrow \text{stochastic function}$$

$$\Rightarrow \hat{\theta}_{ML_{\sim po}} = \underset{\theta}{\arg\max}\ \underset{\underline{x}_m}{E} \left\{ \sum_{q=1}^{Q} \ln^{f(\underline{x}^q|\theta)} + \ln^{f(\underline{x}_m^q)} \right\}$$

$$\Rightarrow \hat{\theta}_{ML_{\sim po}} = \underset{\theta}{\arg\max} \sum_{q=1}^{Q} \underset{\underline{x}_m}{E} \left\{ \ln^{f(\underline{x}^q|\theta)} + \ln^{f(\underline{x}_m^q)} \right\}$$

$$= \underset{\theta}{\arg\max} \sum_{q=1}^{Q} \int d\underline{x}_m \left[\ln^{f(\underline{x}^q|\theta)} + \ln^{f(\underline{x}_m^q)} \right] \underbrace{f(\underline{x}_m)}_{? \to \text{marginalization}}$$

$$= \underset{\theta}{\arg\max} \sum_{q=1}^{Q} \int d\underline{x}_m \left[\ln^{f(\underline{x}^q|\theta)} + \ln^{f(\underline{x}_m^q)} \right] \times \int d\underline{x}\ f(y|\theta)$$

در حالت کلی مسئله به پیدا کردن $f(\underline{x}_m|\theta)$ در فضای توابع که شرایط pdf را دارند تبدیل شد.

برای ساده سازی مسئله فرض می‌کنیم که $f(\underline{x}_m)$ با پارامترهای محدود θ' معلوم خواهد بود. در این صورت

$$\Theta = \begin{pmatrix} \theta \\ \theta' \end{pmatrix}$$

d)

$$\hat{\Theta}_{\substack{ML\\MPO}} = \underset{\Theta}{\arg\max} \sum_{q=1}^{Q} \int d\underline{x}_m \left[\ln^{f(\underline{x}^q|\Theta)} + \ln^{f(\underline{x}_m^q|\Theta)} \right] f(\underline{x}_m|\Theta)$$

حال از آنجایی که حل مسئله optimization سخت است با استفاده از EM مسئله را حل خواهیم کرد.

$$Q(\Theta|\hat{\Theta}(t), \mathcal{X}) = \sum_{q=1}^{Q} \int d\underline{x}_m \underbrace{\left[\ln^{f(\underline{x}^q|\Theta)} + \ln^{f(\underline{x}_m^q|\Theta)} \right]}_{\text{parameter}} \underbrace{f(\underline{x}_m|\hat{\Theta}(t))}_{\text{number}}$$

E-step: Compute $Q(\Theta|\hat{\Theta}(t), \mathcal{X})$

M-step: $\hat{\Theta}(t+1) = \underset{\Theta}{\arg\max}\, Q(\Theta|\hat{\Theta}(t), \mathcal{X})$

ولی در حالت کلی نمی‌توان این مسئله را به صورت EM حل کرد. برای تبدیل این مسئله به یک EM باید فرض‌هایی که طی حل مسئله انجام دادیم را در نظر بگیریم.

1- $f(y|\Theta) = f(\underline{x}|\Theta)\, f(\underline{x}_m|\Theta)$ فرض استقلال قسمت مشاهده نشده از دیگر ابعاد بردار ویژگی

2- $f(\underline{x}_m|\Theta) = f(\underline{x}_m|\Theta')$ فرض ساختاری در مورد distribution قسمت مشاهده نشده.

$\hat{\Theta}(0) \rightarrow \Theta$ (random), Θ' (use ML on $f(\underline{x}_m|\Theta')$ for complete data)

از طریق ML پارامترهای Θ' را از روی داده‌هایی که $\underline{X}_m$ را به صورت مشاهده شده دارند تعیین کنیم.

سوال 5

Problem 5)

* Forward algorith : $\alpha_t(j) = P(o_1, o_2, \dots o_t, q_t = j \mid \lambda)$

$$\alpha_t(j) = \sum_{i=1}^{N} \alpha_{t-1}(i)\, a_{ij}\, b_j(o_t)$$

the previous forward path probability from the prev time step

The transition probability from previous state q_i to current state q_j

The state observation likelihood of the observation symbol o_t given the current state (j)

$$P(O) = \sum_{j=1}^{N} \alpha_t(j)$$

$$\pi = [0.6 \;,\; 0.4] \qquad A = \begin{bmatrix} 0.7 & 0.5 \\ 0.3 & 0.6 \end{bmatrix}$$

$$B = \begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline 0 & 0.1 & 0.4 & 0.5 \\ 1 & 0.6 & 0.3 & 0.1 \end{array}$$

$S \equiv 0 \quad P \equiv 1 \quad E \equiv 2$

PESESEPEPSPEPS $\equiv$ 1202021 2 1012 10

SESPSEPSP SESPE $\equiv$ 0201 0210 1020 12

$P(PESESEPEPSPEPS) = 2.8484 \times 10^{-7}$

$P(SESPSEPSPSESPE) = 1.2124 \times 10^{-7}$

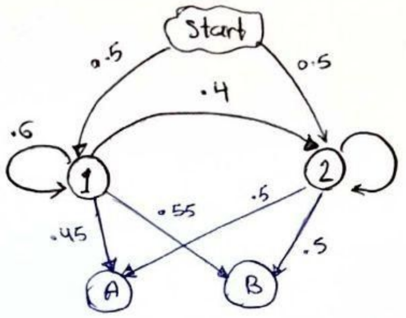
سوال 6

محاسبات احتمال این قسمت و پیاده سازی الگوریتم در فایل hmm.py قرار دارد و نتایج آن در زیر آورده شده است

Problem 5)

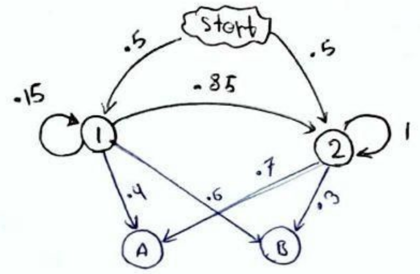
II.

model (1)



$p(ABA) = 0.119897$

model 2



$p(ABA) = 0.120989$

$$p(ABA|\lambda_1) > p(ABA|\lambda_2)$$

I. $p(q_t|q_{t+1},\dots,q_T) \overset{?}{=} p(q_t|q_{t+1})$

$$\frac{p(q_{t+1},q_{t+2},\dots,q_T|q_t)\,p(q_t)}{p(q_{t+1},q_{t+2},\dots,q_T)} \overset{?}{=} \frac{p(q_{t+1}|q_t)\,p(q_t)}{p(q_{t+1})}$$

chain rule

$$\frac{p(q_{t+1}|q_t)\,p(q_{t+2}|q_{t+1},q_t)\cdots p(q_T|q_{T-1},\dots,q_{t+1})}{p(q_{t+1})\,p(q_{t+2}|q_{t+1})\cdots p(q_T|q_{T-1})} \overset{\checkmark}{=} \frac{p(q_{t+1}|q_{(t)})}{p(q_{t+1})}$$

پیوست 1: روند اجرای برنامه

برای سوال ۲ فایل jupyter notebook را باید باز کنید. این فایل نیاز به اجرا ندارد و خروجی در این فایل ذخیره شده است.

برای سوال ۶ پیاده سازی hidden markov model در فایل hmm.py آمده است.