شناسایی آماری الگو

تمرین های سری سه

فرهاد دلیرانی
۹۶۱۳۱۱۲۵

dalirani@aut.ac.ir

dalirani.1373@gmail.com

تمام‌کدها با پایتون 3.6 نوشته شده‌اند.

همچنین از پکیج‌های زیر استفاده کرده ام:

- numpy
- matplotlib

البته برای راحتی در نصب پایتون 3.6 و پکیج های مربوط به دیتاساینس که numpy و matplotlib هم جزیی از آن پکیج‌ها هستند از Anaconda 5.0.0 استفاده کرده‌ام که همه‌ی موارد گفته شده را بدون دردسر و سختی نصب می کند. تنها کافی است آن را از https://www.anaconda.com/download دانلود کنید و Installer باقی کار را انجام می‌دهد.

زبان برنامه نویسی: پایتون 3.6
پکیج‌ها: پکیج‌های گفته شده را برای راحتی در نصب با Anaconda نصب کردم.
ورژن Anaconda من: Anaconda 5.0.0 For Linux Installer که البته همین ورژن برای سایر سیستم عامل‌ها هم موجود است.
محیط برنامه نویسی: pyCharm Community Edition

سوال ۱)

از آنجایی که Ln صعودی است و در میزان تتا تاثیر ندارد از likelihood لگاریتم طبیعی گرفته‌ام.

1)

a)

$$f(x_k;\theta)=\frac{x_k}{\theta^2}\exp\left(-\frac{x_k^2}{2\theta^2}\right) \qquad x_k\geq 0 \qquad \theta>0$$

$$L(\theta)=\ln p(D|\theta)=\ln\prod_{k=1}^{N} f(x_k;\theta)$$

$$=\ln\prod_{k=1}^{N}\frac{x_k}{\theta^2}\exp\left(-\frac{x_k^2}{2\theta^2}\right)=\sum_{k=1}^{N}\left(\ln\frac{x_k}{\theta^2}+\ln\exp\left(-\frac{x_k^2}{2\theta^2}\right)\right)$$

$$=\sum_{k=1}^{N}\left(\ln\frac{x_k}{\theta^2}-\frac{x_k^2}{2\theta^2}\right)\Rightarrow$$

$$\frac{L(\theta)}{d\theta}=\frac{d}{d\theta}\sum_{k=1}^{N}\left(\ln x_k-2\ln\theta-\frac{x_k^2}{2\theta^2}\right)$$

$$=\sum_{k=1}^{N}\left(-\frac{2}{\theta}-\frac{x_k^2}{2}\left(\frac{0-2\theta}{\theta^4}\right)\right)$$

$$=\sum_{k=1}^{N}\left(-\frac{2}{\theta}-\frac{x_k^2}{2}\left(-\frac{2}{\theta^3}\right)\right)=\sum_{k=1}^{N}\left(-\frac{2}{\theta}+\frac{x_k^2}{\theta^3}\right)$$

$$\to\frac{L(\theta)}{d\theta}=0\to\sum_{k=1}^{N}\left(-\frac{2}{\theta}+\frac{x_k^2}{\theta^3}\right)=0\to-\frac{2N}{\theta}+\frac{1}{\theta^3}\sum_{k=1}^{N}x_k^2=0$$

$$\to\frac{2N}{\theta}=\frac{1}{\theta^3}\sum_{k=1}^{N}x_k^2\to\theta^2=\frac{1}{2N}\sum_{k=1}^{N}x_k^2$$

$$\boxed{\theta=\sqrt{\frac{1}{2N}\sum_{k=1}^{N}x_k^2}}$$

1)

b) $f(x_k;\theta) = \sqrt{\theta}\; x_k^{\sqrt{\theta}-1}$ $\qquad \phi \leq x_k \leq 1 \qquad \theta > \phi$

$$l(\theta) = \ln \prod_{k=1}^{N} \sqrt{\theta}\; x_k^{\sqrt{\theta}-1}$$

$$= \sum_{k=1}^{N} \left(\frac{1}{2} \ln \theta + (\sqrt{\theta}-1) \ln x_k \right)$$

$$\frac{l(\theta)}{d\theta} = \sum_{k=1}^{N} \left(\frac{1}{2\theta} + \frac{\ln x_k}{2\sqrt{\theta}} \right)$$

$$\frac{L(\theta)}{d\theta} = \phi \longrightarrow \sum_{k=1}^{N} \left(\frac{1}{2\theta} + \frac{\ln x_k}{2\sqrt{\theta}} \right) = \phi$$

$$\longrightarrow \frac{1}{2\theta} \sum_{k=1}^{N} 1 + \frac{1}{2\sqrt{\theta}} \sum_{k=1}^{N} \ln x_k = \phi \longrightarrow$$

$$\longrightarrow \frac{N}{2\theta} + \frac{1}{2\sqrt{\theta}} \sum_{k=1}^{N} \ln x_k = 0$$

$$\longrightarrow \frac{N + \sqrt{\theta} * \sum_{k=1}^{N} \ln x_k}{2\sqrt{\theta}\sqrt{\theta}} = 0$$

$$\longrightarrow \sqrt{\theta} = \frac{-N}{\sum_{k=1}^{N} \ln x_k} \rightarrow \boxed{\theta = \left(\frac{-N}{\sum_{k=1}^{N} \ln x_k} \right)^2}$$

سوال ۲)

از آنجایی که Ln صعودی است و در میزان تتا تاثیر ندارد از likelihood لگاریتم گرفته‌ام.

2)

a)

$\cancel{0} \leq X_1 \leq X_2 \leq \cdots \leq X_n \leq \theta$

$$L(\theta) = \ln p(D|\theta) = \ln \prod_{k=1}^{n} \frac{1}{\theta} = \ln \theta^{-n} = -n \ln \theta$$

$$\frac{dL(\theta)}{d\theta} = -\frac{n}{\theta}$$

مشتق برای $\theta > 0$ ، منفی است. ولی از آن جایی که $L(\theta)$ نزولی است

هرچه θ بزرگ تری انتخاب کنیم مقدار تابع likelihood ما کم تر می شود؛ در نتیجه باید

کوچک ترین مقدار ممکن را برای θ انتخاب کنیم و از آن جایی که $x_n \leq \theta$

است، کوچک ترین مقدار ممکن برای θ همان x_n می شود و تابع likelihood

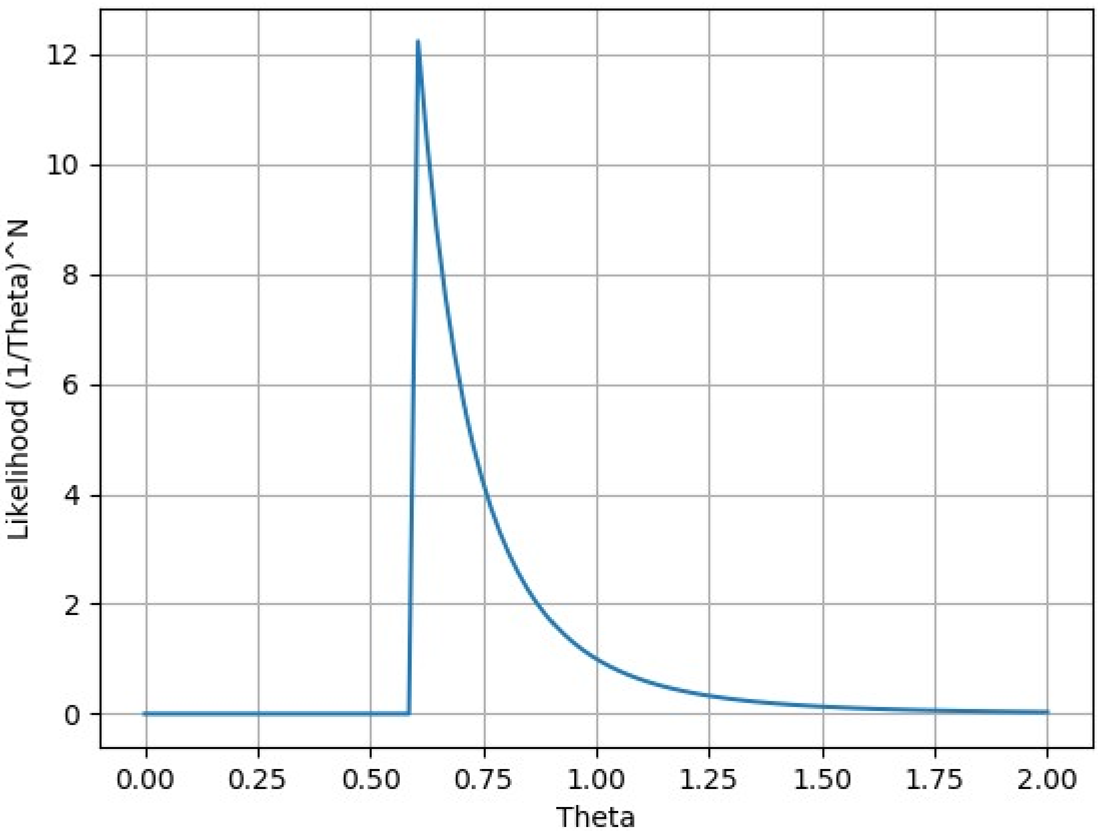
در آن ماکزیمم می شود

$$\boxed{\theta = x_n}$$

بخش b)

برای رسم نمودار کدی نوشته‌ام که در پوشه‌ی problem2 و در فایل problem2.py موجود است.

خروجی حاصل از اجرای کد را در تصویر زیر مشاهده می‌کنید:

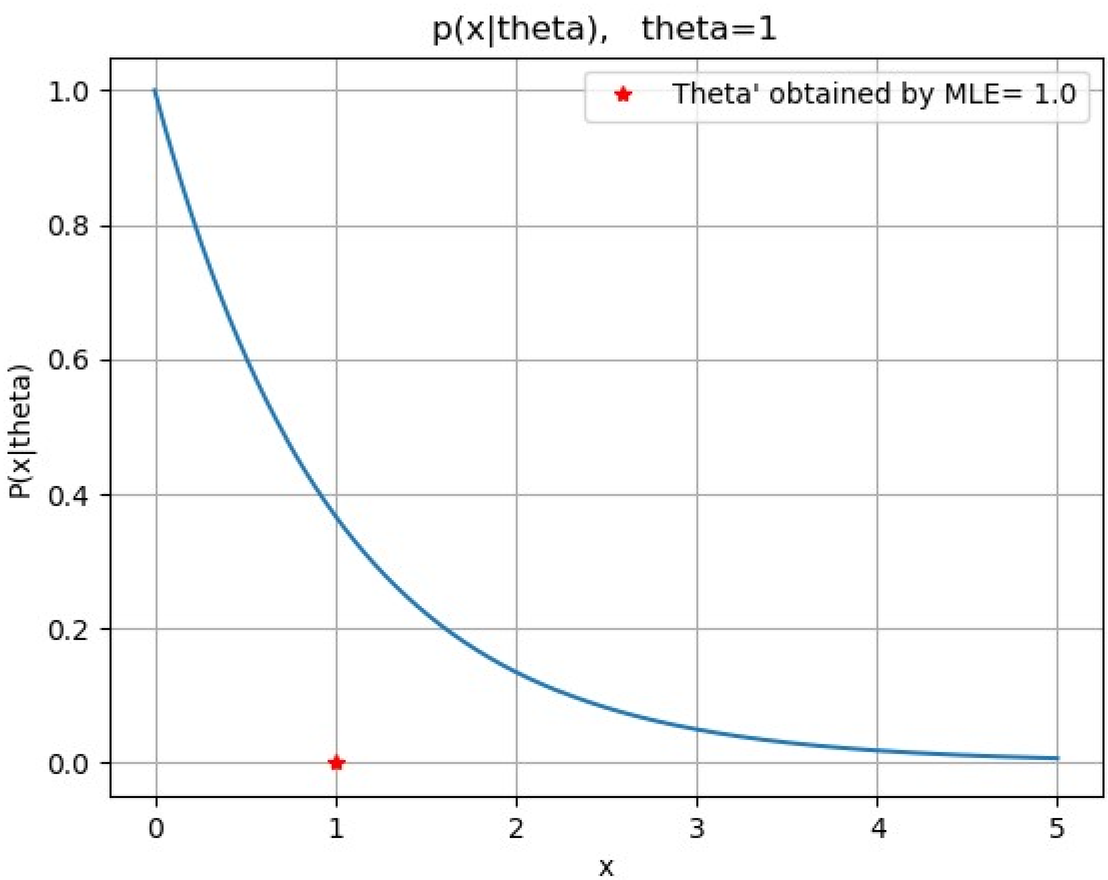


سوال ۳) q1

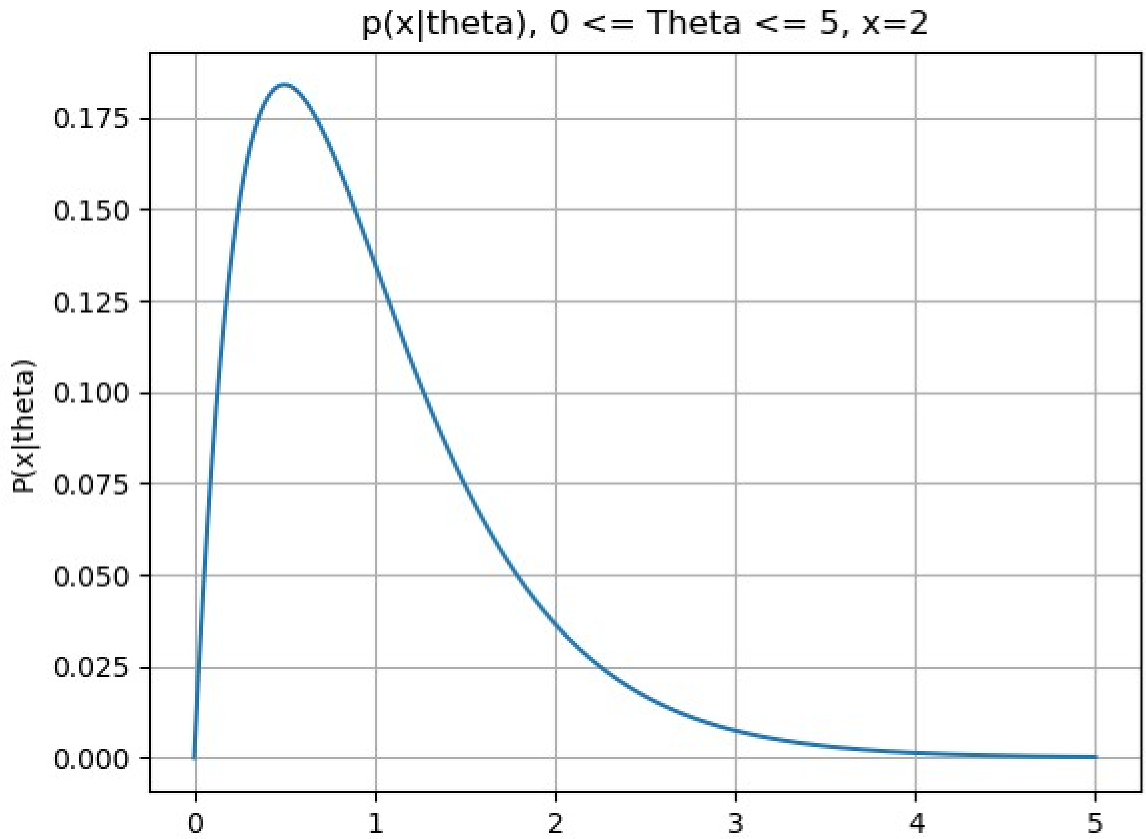
بخش A و C)

برای انجام بهش A و C کد پایتونی نوشته‌ام که در پوشه‌ی problem3 و در فایل problem3-q1.py قرار دارد. خروجی حاصل از اجرای کد را در دو تصویر زیر مشاهده می‌کنید:

مقدار P(x| theta) به ازای تتا مساوی یک، همچنین مقدار تتایی که با MLE در بخش C این سوال خواسته شده است را با * در تصویر زیر مشخص کرده‌ام که مقدارش برابر یک است.



مقدار p(x| theta) برای x=2 و تتاهای بین 0 و n را در تصویر زیر مشاهده می‌کنید:


p(x|theta), 0 <= Theta <= 5, x=2
P(x|theta)
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
0
1
2
3
4
5

بخش B)

3) Q1) B)

$$L(\theta) = \ln \prod_{k=1}^{n} \theta e^{-\theta x_k} = \sum_{k=1}^{n} \left(\ln\theta + \ln e^{-\theta x_k} \right)$$

$$= \sum_{k=1}^{n} \left(\ln\theta - \theta x_k \right)$$

$$\rightarrow \frac{d L(\theta)}{d\theta} = \sum_{k=1}^{n} \left(\frac{1}{\theta} - x_k \right) = 0$$

$$\rightarrow \frac{n}{\theta} - \sum_{k=1}^{n} x_k = 0 \rightarrow \frac{1}{\theta} = \frac{1}{n} \sum_{k=1}^{n} x_k$$

$$\rightarrow \boxed{\theta = \frac{1}{\frac{1}{n} \sum_{k=1}^{n} x_k}}$$

بخش C)

پاسخ به این بخش در بخش A همین سوال داده شده است. مقدار تتا که با MLE به دست می آید برابر با یک است.

سوال ۳ – q3)

3) Q3)

a)

براساس سوال:

$$p(z_{ik}=1 \mid p(\omega_i)) = p(\omega_i) = p(\omega_i)^1 (1-p(\omega_i))^0$$

$$p(z_{ik}=0 \mid p(\omega_i)) = 1-p(\omega_i) = p(\omega_i)^0 (1-p(\omega_i))^1$$

که اگر دو عبارت بالا را به شکل زیر ترکیب کنیم دیگر لازم به ذکر مقدار z_{ik} نیستیم

$$p(z_{ik} \mid p(\omega_i)) = p(\omega_i)^{z_{ik}} (1-p(\omega_i))^{1-z_{ik}}$$

و از آن جایی که نمونه‌ها مستقل هستند:

$$p(z_{i1}, z_{i2}, \ldots, z_{in} \mid p(\omega_i)) = p(z_{i1} \mid p(\omega_i))\, p(z_{i2} \mid p(\omega_i)) \cdots p(z_{in} \mid p(\omega_i))$$

$$= \prod_{k=1}^{n} p(\omega_i)^{z_{ik}} (1-p(\omega_i))^{1-z_{ik}}$$

b) $\theta = p(\omega_i)$

$$L(\theta) = \ln \prod_{k=1}^{n} p(\omega_i)^{z_{ik}} (1-p(\omega_i))^{1-z_{ik}}$$

$$= \sum_{k=1}^{n} \left(z_{ik} \ln p(\omega_i) + (1-z_{ik}) \ln (1-p(\omega_i)) \right)$$

$$\rightarrow \frac{dL(\theta)}{d\theta} = \sum_{k=1}^{n} \left(\frac{z_{ik}}{p(\omega_i)} + \frac{(1-z_{ik})(-1)}{1-p(\omega_i)} \right) = 0$$

$$\rightarrow \sum_{k=1}^{n} \left(\frac{z_{ik}(1-p(\omega_i)) + (z_{ik}-1)p(\omega_i)}{p(\omega_i)(1-p(\omega_i))} \right) = 0 \rightarrow$$

$$\rightarrow \sum_{k=1}^{n} \left(\frac{z_{ik} - z_{ik}p(\omega_i) + z_{ik}p(\omega_i) - p(\omega_i)}{p(\omega_i)(1-p(\omega_i))} \right) = 0 \rightarrow \sum_{k=1}^{n} (z_{ik} - p(\omega_i)) = 0$$

$$\Rightarrow \boxed{\hat{p}(\omega_i) = \frac{1}{n} \sum_{k=1}^{n} z_{ik}}$$

سوال ۳ – q4)

4)

$$p(x|\theta) = \prod_{i=1}^{d} \theta_i^{x_i} (1-\theta_i)^{1-x_i}$$

$$p(x_1, x_2, \dots, x_n|\theta) = \prod_{i=1}^{n} \prod_{j=1}^{d} \theta_j^{x_{ij}} (1-\theta_j)^{1-x_{ij}}$$

$$L(\theta) = \ln \prod_{i=1}^{n} \prod_{j=1}^{d} \theta_j^{x_{ij}} (1-\theta_j)^{1-x_{ij}}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{d} \left(x_{ij} \ln\theta_j + (1-x_{ij}) \ln(1-\theta_j) \right)$$

$$\left[\nabla_\theta L(\theta)\right]_j = \sum_{i=1}^{n} \left(\frac{x_{ij}}{\theta_j} - \frac{(1-x_{ij})}{1-\theta_j} \right)$$

$$= \frac{1}{\theta_j} \sum_{i=1}^{n} x_{ij} - \frac{1}{1-\theta_j} \sum_{i=1}^{n} (1-x_{ij}) = 0$$

$$\rightarrow (1-\theta_j) \sum_{i=1}^{n} x_{ij} = \theta_j \sum_{i=1}^{n} (1-x_{ij})$$

$$\rightarrow (1-\theta_j) \sum_{i=1}^{n} x_{ij} = \theta_j n - \theta_j \sum_{i=1}^{n} x_{ij}$$

$$\rightarrow \theta_j n = \sum_{i=1}^{n} x_{ij} \longrightarrow \theta_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij}$$

$$\longrightarrow \boxed{\theta = \frac{1}{n} \sum_{i=1}^{n} x_i}$$

سوال ۳ – q5)

قسمت a)

5)

a)

$$p(X|w_1)=\prod_{i=1}^{d}\rho^{x_i}(1-\rho)^{1-x_i}$$

$\theta=\rho$

$$l(\theta)=\ln\prod_{i=1}^{d}\rho^{x_i}(1-\rho)^{1-x_i}=\sum_{i=1}^{d}\left(x_i\ln\rho+(1-x_i)\ln(1-\rho)\right)\Rightarrow$$

$$\frac{d\,l(\theta)}{d\theta}=\sum_{i=1}^{d}\left(\frac{x_i}{\rho}+\frac{(1-x_i)(-1)}{(1-\rho)}\right)=\sum_{i=1}^{d}\left(\frac{x_i}{\rho}+\frac{(x_i-1)}{1-\rho}\right)=0\rightarrow$$

$$\frac{1}{\rho}\sum_{i=1}^{d}x_i+\frac{1}{1-\rho}\sum_{i=1}^{d}(x_i-1)=0\rightarrow$$

$$\frac{1}{\rho}\sum_{i=1}^{d}x_i+\frac{1}{1-\rho}\sum_{i=1}^{d}x_i-\frac{d}{1-\rho}=0\rightarrow$$

$$\left(\sum_{i=1}^{d}x_i\right)\left(\frac{1}{\rho}+\frac{1}{1-\rho}\right)-\frac{d}{1-\rho}=0\longrightarrow\left(\sum_{i=1}^{d}x_i\right)\frac{1-\cancel{\rho}+\cancel{\rho}}{(\rho)(1-\rho)}-\frac{d}{1-\rho}=0$$

$$\rightarrow\left(\sum_{i=1}^{d}x_i\right)=\frac{d}{(1-\rho)}*(1-\rho)(\rho)=\rho d\rightarrow\boxed{\rho=\frac{1}{d}\sum_{i=1}^{d}x_i}$$

قسمت b)

5)

b)

$$P(x_i|w_1) = P^{x_i}(1-P)^{1-x_i}$$

$$E[x_i] = 0 * P^0(1-P)^1 + 1 * P^1(1-P)^0 = P$$

$$E\left[\frac{1}{d}\sum_{i=1}^{d} x_i\right] = \frac{d}{d}P = P$$

$$E[x_i^2] = 0^2 * P^0 * (1-P)^1 + 1^2 * P^1 * (1-P)^0 = P$$

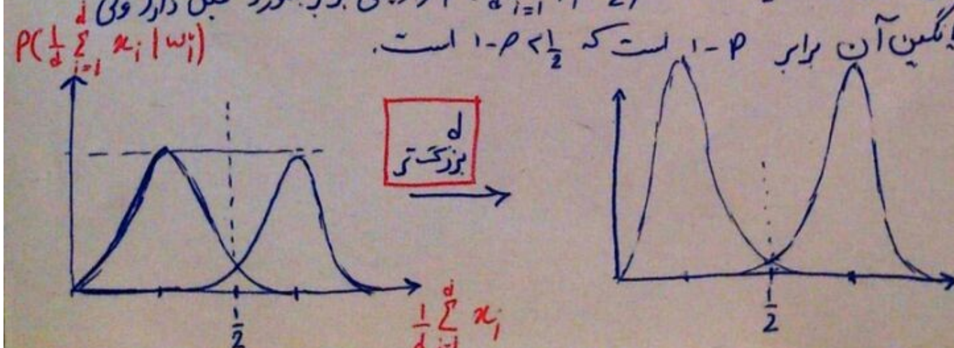
$$\rightarrow Var[x_i] = E[x_i^2] - (E[x_i])^2 = P - P^2$$

$$Var\left[\frac{1}{d}\sum_{i=1}^{d} x_i\right] = \frac{1}{d^2} Var\left[\sum_{i=1}^{d} x_i\right] = \frac{1}{d^2}\left[d * (P - P^2)\right]$$

$$= \frac{P(1-P)}{d}$$

در $d \to \infty$ مقدار واریانس صفر می‌شود.

* پس اگر $P(\frac{1}{d}\sum_{i=1}^{d} x_i|w_1)$ را با توزیع نرمال نشان دهیم میانگین آن برابر با P می‌شود و واریانس آن برابر با $\frac{P(1-P)}{d}$ که در d های بزرگ صفر می‌شود و طبق سوال $\frac{1}{2} < P$ است. $P(\frac{1}{d}\sum_{i=1}^{d} x_i|w_2)$ واریانس برابر با مورد قبل دارد ولی میانگین آن برابر $1-P$ است که $1-P < \frac{1}{2}$ است.



b) b)

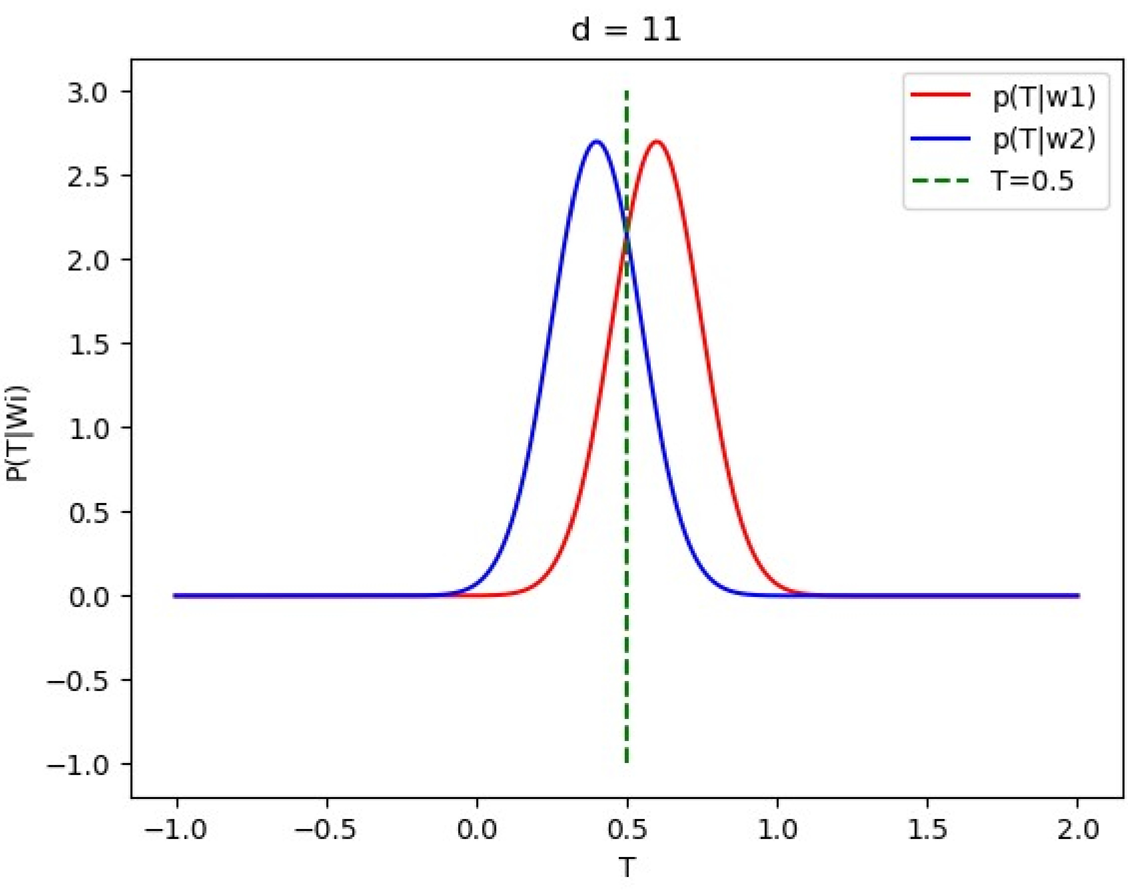
ادامه)

هرچه d بزرگ تر می‌شود واریانس کم تری می‌شود و اشتراک
دو توزیع کم تری می‌شود در نتیجه خطا به صفر میل می‌کند.

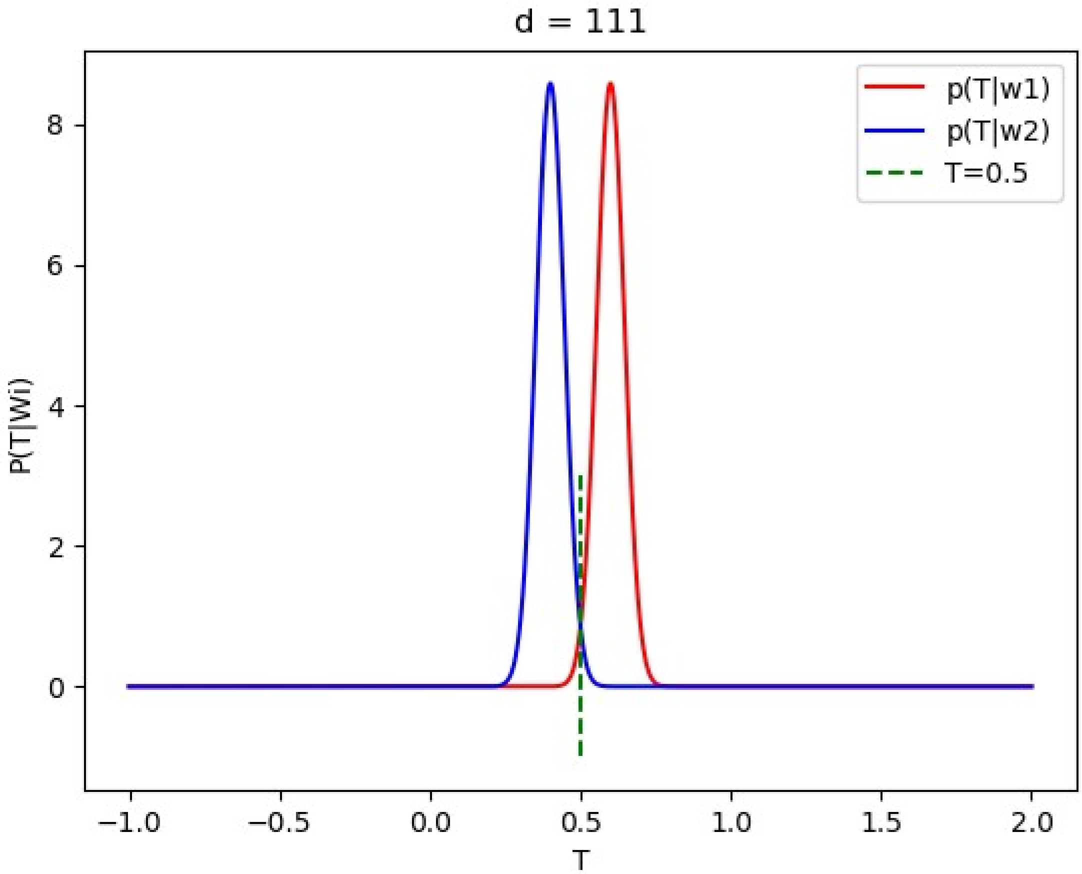
قسمت C)

برای این قسمت از سوال کد پایتونی نوشته‌ام که در پوشه‌ی problem3 و در فایل problem3-q5.py قرار دارد و خروجی‌های حاصل از اجرای آن را در زیر مشاهده می‌کنید:

P(T| Wi) به ازای d = 11:



P(T| Wi) به ازای d = 111:



همین طور که مشاهده می‌کنید با زیاد شدن میزان d واریانس که از رابطه‌ی p(1-p)/d حساب می‌شود کوچک و کوچک‌تر می‌شود. و ناحیه‌ی مشترک بین توزیع نرمال‌ها کمتر می‌شود و در نتیجه خطا کمتر می‌شود.

Computer Exercise 1)

کدهای این سوال در پوشه‌ی ComputerExercises-1 و در فایل computerExercise-1.py قرار دارد.

بخش A)

میانگین و واریانس برای هر فیچر w1:

```
Part A:
Mean feature0 =[-0.07089999999999998]
Mean feature1 =[-0.6047000000000001]
Mean feature2 =[-0.9109999999999999]

Var feature0 =[[ 0.90617729]]
Var feature1 =[[ 4.20071481]]
Var feature2 =[[ 4.541949]]
```

بخش B)

میانگین و واریانس هر دو فیچر ممکن w1:

```
Part B:
Mean feature01 =[-0.07089999999999998, -0.6047000000000001]
Mean feature02 =[-0.07089999999999998, -0.9109999999999999]
Mean feature12 =[-0.6047000000000001, -0.9109999999999999]

Covariance Matrix feature01 =
[[ 0.90617729  0.56778177]
 [ 0.56778177  4.20071481]]
Covariance Matrix feature02 =
[[ 0.90617729  0.3940801 ]
 [ 0.3940801   4.541949  ]]
Covariance Matrix feature12 =
[[ 4.20071481  0.7337023 ]
 [ 0.7337023   4.541949  ]]
```

بخش C)

میانگین و واریانس هر سه فیچر w1:

```
Part C:
Mean feature012 =[-0.07089999999999998, -0.6047000000000001, -0.9109999999999999]

Covariance Matrix feature012 =
[[ 0.90617729  0.56778177  0.3940801 ]
 [ 0.56778177  4.20071481  0.7337023 ]
 [ 0.3940801   0.7337023   4.541949  ]]
```

بخش D)

میانگین و واریانس هر سه فیچر w2 با فرض قطری بودن ماتریس کواریانس و همین طوری میانگین و واریانس هر فیچر به صورت جدا از هم:

```
Part D:
Mean feature0 =[-0.11259999999999999]
Mean feature1 =[0.42989999999999995]
Mean feature2 =[0.003720000000000001]

Variance Matrix feature0 =[[ 0.05392584]]
Covariance Matrix feature1 =[[ 0.04597009]]
Covariance Matrix feature2 =[[ 0.00726551]]

Mean feature012 =[-0.11259999999999999, 0.42989999999999995, 0.003720000000000001]

Covariance feature012(assumed diagonal) =
[[ 0.05392584  0.          0.        ]
 [ 0.          0.04597009  0.        ]
 [ 0.          0.          0.00726551]]


Process finished with exit code 0
```

بخش e و f)

همین طور که در نتایج بخش‌های مختلف این سوال مشاهده می‌کنید، فرقی نمی‌کند میانگین را برای هر فیچر جدا حساب کنیم و یا برای چند فیچر. میانگین متناظر با هر ستون در همه‌ی روش‌ها برابر هستند.

همین طور که در نتایج بخش‌های مختلف این سوال مشاهده می‌کنید، فرقی نمی‌کند واریانس را برای هر فیچر جدا حساب کنیم و یا برای چند فیچر ماتریس کواریانس را حساب کنیم. واریانس متناظر با فیچر در همه‌ی روش‌ها برابر هستند.

Computer Exercise 2: در متن سوال P(Theta) در کتابی که بر روی LMS است داده نشده است و چند نسخه‌ی دیگر از کتاب را که در اینترنت هست را دیدم و در آن‌ها هم موجود نبود. معمولا نسخه‌های موجود در اینترنت نسخه‌ای است که پیش از چاپ نسخه‌ی اصلی کتاب در دسترس قرار گرفته اند.

computer exercise 3:

کدهای این بخش در پوشه‌ی computerExercises-3 و در فایل computerExercise-3.py موجود هستند.

بخش A)

برای این بخش تابع `p_x_given_D` را نوشته ام که با دریافت `mu0, sigma0, sigma, D` مقدارهای *`p(x|D)~N(muN, sigma+sigmaN)`* را محاسبه و رسم می‌کند.

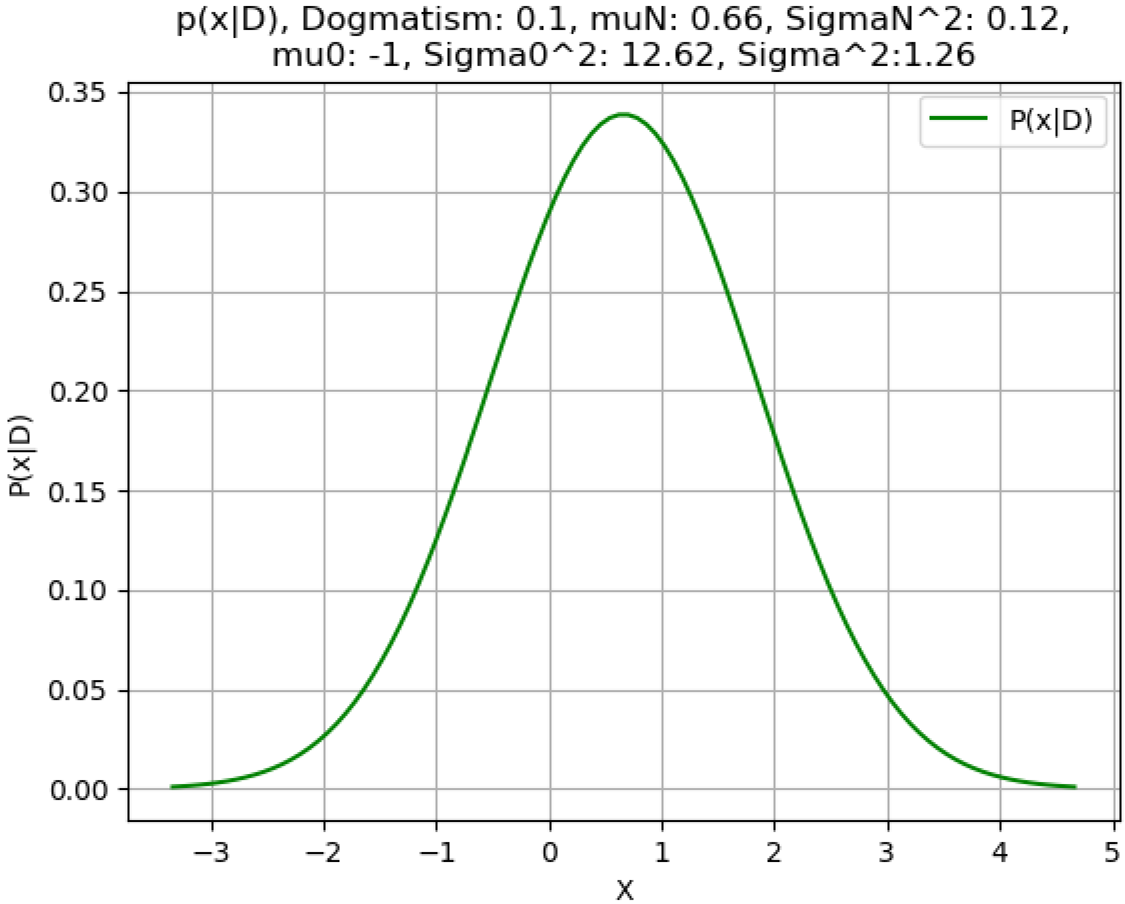
بخش B)

برای این بخش تابع `p_x_given_D` را برای پارامترهای گفته شده فراخوانده‌ام که خروجی اجرای کد را در تصویرهای زیر مشاهده می‌کنید:

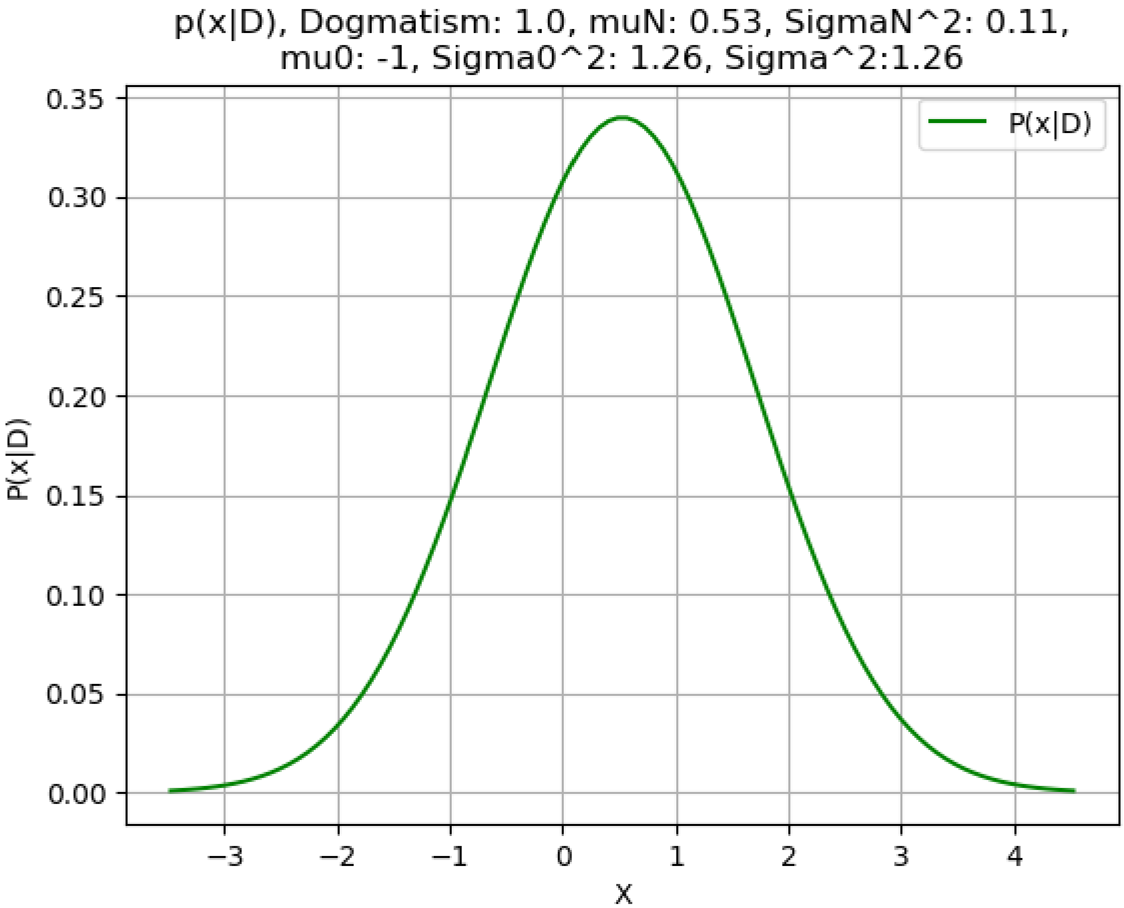
واریانس و میانگین ویژگی x2 داده‌های w3 :

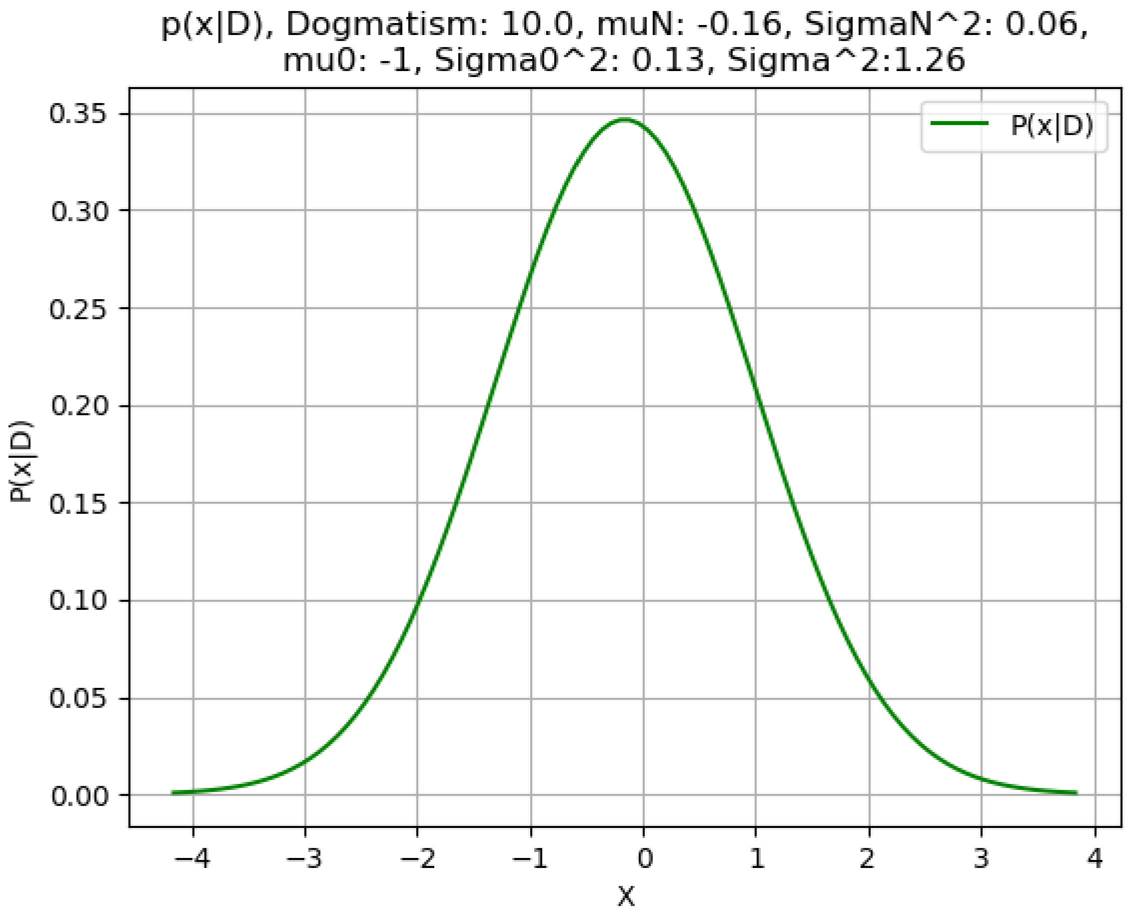
```
part-B)
Estimated Mean for x2 of w3:
 0.6786
Estimated Sigma for x2 of w3:
 1.1234507732873746
Estimated variance(Sigma^2) for x2 of w3:
 1.26214164
```

Dogmatism 0.1

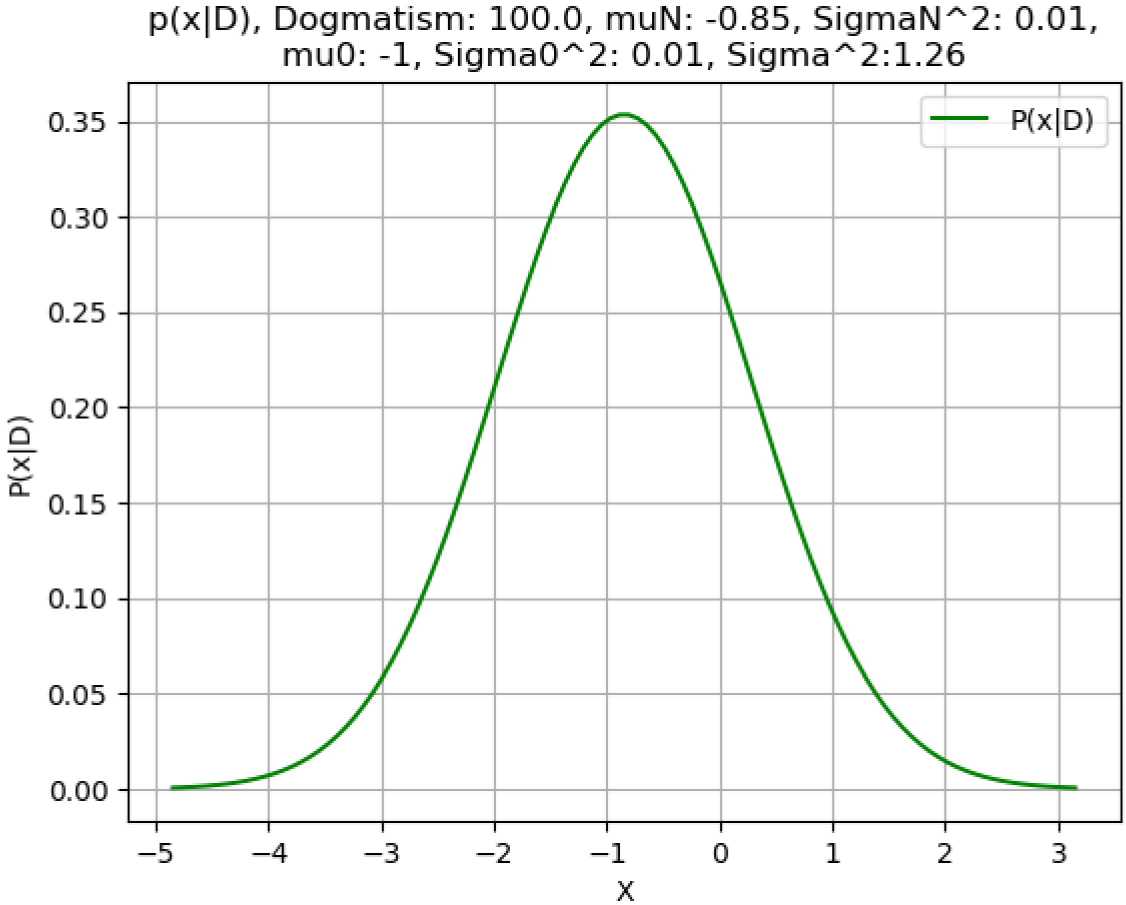

p(x|D), Dogmatism: 0.1, muN: 0.66, SigmaN^2: 0.12,
mu0: -1, Sigma0^2: 12.62, Sigma^2:1.26
P(x|D)
P(x|D)
X
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−3
−2
−1
0
1
2
3
4
5

Dogmatism 1


p(x|D), Dogmatism: 1.0, muN: 0.53, SigmaN^2: 0.11,
mu0: -1, Sigma0^2: 1.26, Sigma^2:1.26
P(x|D)
P(x|D)
X
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−3
−2
−1
0
1
2
3
4



Dogmatism 100


p(x|D), Dogmatism: 100.0, muN: -0.85, SigmaN^2: 0.01,
mu0: -1, Sigma0^2: 0.01, Sigma^2:1.26
P(x|D)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
P(x|D)
−5
−4
−3
−2
−1
0
1
2
3
X