دانشگاه تهران

پردیس دانشکده‌های فنی

دانشکده برق و کامپیوتر

گزارش تمرین شماره 5

درس سیستم های هوشمند

پاییز 1401

امیرحسین بیرژندی

...

810198367

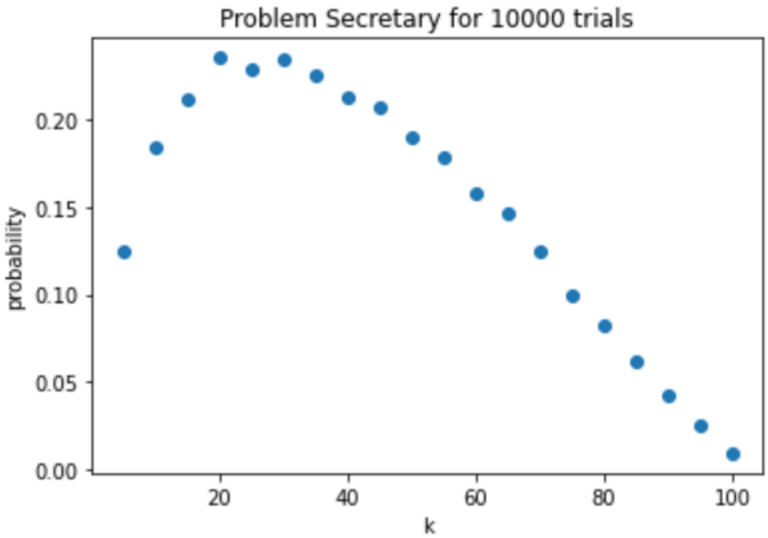
...

سوال 1 – سوالات آماری (پیاده‌سازی)

بخش 1

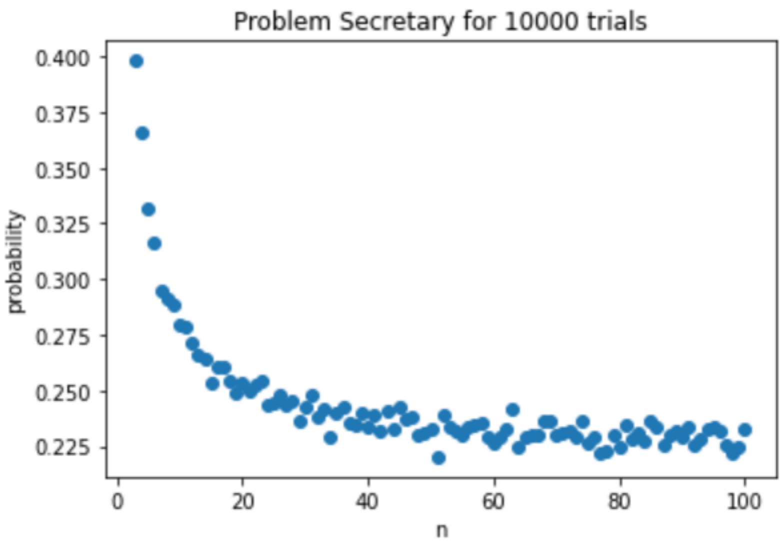
.1

```
best k:20
```



بهترین k در این سری آزمایش 20 بدست آمده است؛ اما توجه شود که ممکن است در یک سری آزمایش دیگر k اپتیمم برابر مقدار دیگری در همین حوالی باشد.

.2



.3

$$P(r) = \sum_{i=1}^{n} P(\text{applicant } i \text{ is selected} \cap \text{applicant } i \text{ is the best})$$

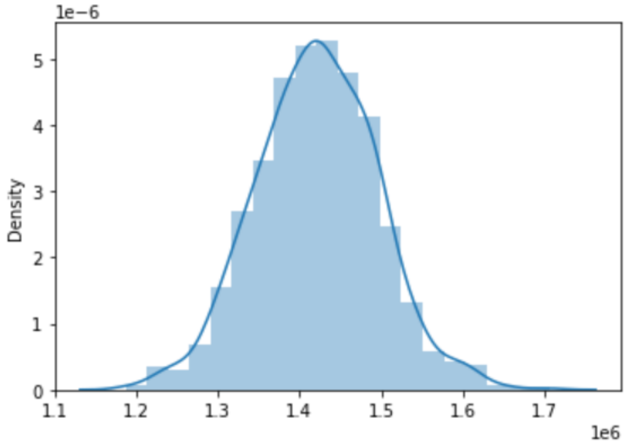
$$= \sum_{i=1}^{n} P(\text{applicant } i \text{ is selected} \mid \text{applicant } i \text{ is the best})$$

$$= \left[\sum_{i=1}^{r-1} 0 + \sum_{i=r}^{n} P\left(\begin{array}{c} \text{the best of the first} \\ i-1 \text{ applicants is in the} \\ \text{first } r-1 \text{ applicants} \end{array} \middle| \text{ applicant } i \text{ is the best} \right) \right] \cdot \frac{1}{n}$$

$$= \left[\sum_{i=r}^{n} \frac{r-1}{i-1} \right] \cdot \frac{1}{n} = \frac{r-1}{n} \sum_{i=r}^{n} \frac{1}{i-1}$$

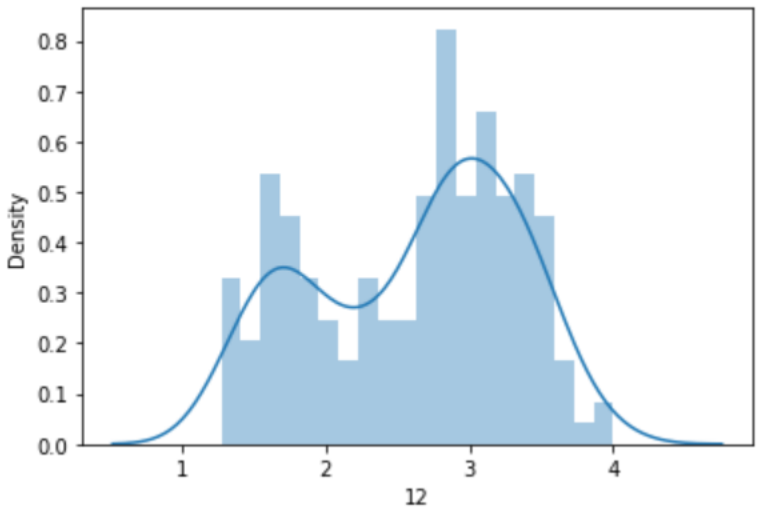
بخش 2

برای بخش اول این سوال از n = 1000 و s = 1000 استفاده می‌کنیم. نمودار توزیع میانگین این نمونه‌ها به شرح زیر است:



حال نمودار توزیع داده های ستون 12 دیتاست ”wine“ را در شکل زیر رسم می‌کنیم.

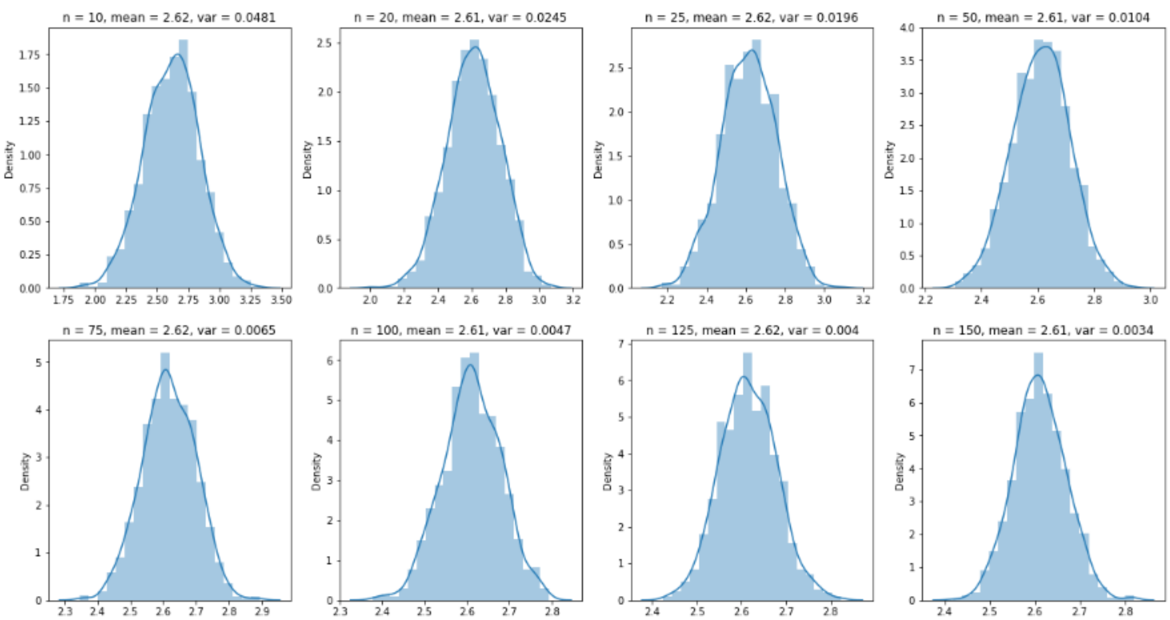
```
mean:  2.6116853932584267
variance:  0.501254462820351
```



حال برای این بخش به ازای n های مختلف نمونه های متفاوتی گرفته و میانگین و واریانس میانگین این نمونه ها را محاسبه می‌کنیم. طبق قضییه حد مرکزی میانگین این نمونه ها باید تقریبا برابر میانگین اصلی μ و واریانس این نمونه ها برابر $\frac{\sigma^2}{n}$ باشد. در نتیجه توقع داریم با افزایش n مقدار واریانس داده ها کاهش یابد که این موضوع را در نتایج بالا می‌توانیم نظاره‌گر باشیم.

```
n = 10
clt mean = 2.61
clt variance = 0.0501
mean = 2.62
variance = 0.0481

n = 20
clt mean = 2.61
clt variance = 0.0251
mean = 2.61
variance = 0.0245

n = 25
clt mean = 2.61
clt variance = 0.0201
mean = 2.62
variance = 0.0196

n = 50
clt mean = 2.61
clt variance = 0.01
mean = 2.61
variance = 0.0104

n = 75
clt mean = 2.61
clt variance = 0.0067
mean = 2.62
variance = 0.0065

n = 100
clt mean = 2.61
clt variance = 0.005
mean = 2.61
variance = 0.0047

n = 125
clt mean = 2.61
clt variance = 0.004
mean = 2.62
variance = 0.004

n = 150
clt mean = 2.61
clt variance = 0.0033
mean = 2.61
variance = 0.0034
```

مشاهده می‌کنیم مقادیر به دست آمده از روی نمونه ها با مقدار مورد انتظار در قضییه حد مرکزی یکسان است.



سوال 2 – سوالات آماری (تحلیلی)

بخش 1 :

independence: $f_{XY}(x,y) = f_X(x) f_Y(y) \rightarrow$ X and Y are independent

uncorrelation: $$\rho = E\left[\frac{(X-E(X))}{\sqrt{E[(X-E(X))^2]}} \frac{(Y-E(Y))}{\sqrt{E[(Y-E(Y))^2]}}\right] = 0$$

X and Y are uncorrelated

* حال در آمار و احتمال به راحتی می‌توان نشان داد که اگر دو متغیر مستقل باشند، ناهمبسته هم می‌باشند
از آن طرف ناهمبستگی لزوماً استقلال را نتیجه نمی‌دهد.

دلیل :

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var[X]Var[Y]}}$$

$$Cov(X,Y) = E[XY] - E[X]E[Y] \quad \text{(II)}$$

استقلال $\Rightarrow E[XY] = E[X]E[Y] \overset{\text{(II)}}{\Rightarrow} Cov(X,Y) = 0 \Rightarrow \rho = 0$ ✓

ناهمبستگی $\Rightarrow \rho = 0 \Rightarrow Cov(X,Y) = 0 \Rightarrow$ لزوماً مستقل نیستند با مثال نقض

* حال از آنجایی که احمد داده‌ها را به ابعاد جدیدی منتقل کرده که داده‌ها در آنجا ناهمبسته هستند
نمی‌توان لزوماً نتیجه گرفت که داده‌ها مستقل نیز می‌باشند. در نتیجه استدلال احمد صحیح نیست.

سوال ۲ - بخش ۲

تابع چگالی احتمال :

$$f_X(x_j) = (2\pi\sigma_0^2)^{-1/2} \exp\left(\frac{-1}{2} \frac{(x_j - \mu_0)^2}{\sigma_0^2} \right)$$

حال تابع likelihood را اینگونه تعریف می‌کنیم :

$$L(\mu, \sigma^2, x_1, \dots, x_n) = (2\pi\sigma^2)^{-n/2} \exp\left(\frac{-1}{2\sigma^2} \sum_{j=1}^{n} (x_j - \mu)^2 \right)$$

حال با اخذ لگاریتم این تابع به log-likelihood function می‌رسیم .

$$l(\mu, \sigma^2; x_1, x_2, \dots, x_n) = \frac{-n}{2}\ln(2) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^{n} (x_j - \mu)^2$$

حال سعی بر ماکزیمم کردن log-likelihood function می‌کنیم .

$$\max_{\mu, \sigma^2} l(\mu, \sigma^2; x_1, \dots, x_n)$$

$$\frac{\partial}{\partial \mu} l(\mu, \sigma^2; x_1, \dots, x_n) = 0$$

$$\frac{\partial}{\partial \sigma^2} l(\mu, \sigma^2; x_1, \dots, x_n) = 0$$

$$\frac{\partial}{\partial \mu} l(\mu, \sigma^2; x_1, \dots, x_n) = \frac{\partial}{\partial \mu} \left(\frac{-n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2} \sum (x_j - \mu)^2 \right)$$

$$= \frac{1}{\sigma^2} \sum_{j=1}^{n} (x_j - \mu) = \frac{1}{\sigma^2} \left(\sum_{j=1}^{n} x_j - n\mu \right)$$

$$\Rightarrow \sum_{j=1}^{n} x_j - n\mu = 0 \Rightarrow \mu = \frac{1}{n} \sum_{j=1}^{n} x_j$$

$$\frac{\partial}{\partial \sigma^2} l(\mu, \sigma^2; x_1, \dots, x_n) = \frac{-n}{2\sigma^2} + \left[\frac{1}{2} \sum_{j=1}^{n} (x_j - \mu)^2 \right] \frac{1}{(\sigma^2)^2}$$

$$= \frac{1}{2\sigma^2} \left[\frac{1}{\sigma^2} \sum_{j=1}^{n} (x_j - \mu)^2 - n \right] = 0$$

$$\Rightarrow \sigma^2 = \frac{1}{n} \sum_{j=1}^{n} (x_j - \mu)^2$$

سوال 3

.1

با استفاده از 4 ویژگی مدل خود را پیاده‌سازی می‌کنیم که جلوتر آن‌ها را توضیح می‌دهیم.

ارتفاع (Height): ارتفاع می‌تواند ویژگی خوبی برای جداسازی اعداد باشد زیرا عددی مثل 0 و یا 6 معمولا با ارتفاع کمتری نسبت به 7 و یا 8 نوشته می‌شود.

تعداد هشتگ ها: می تواند معیار خوبی برای جداسازی باشد زیرا هشتگ ها که ضخامت اعداد را مشخص می‌کنند می‌تواند معیار خوبی برای جداسازی اعدادی مثل 1 و 8 باشد.

تعداد مثبت ها: مثبت ها که محیط یک عدد را مشخص می‌کنند می تواند معیار خوبی برای جداسازی اعدادی مثل 8 و 9 از 7 و 1 باشد.

تعداد کاراکتر های ردیف اول: عددی مانند 1 و یا 8 در ردیف اول کاراکتر های بسیار کمتری نسبت به 7 و 5 دارند در نتیجه معیار خوبی خواهند بود.

شرح الگوریتم:

در این اگوریتم ابتدا

1. احتمال prior که همان درصد تعداد هر لیبل نسبت به کل داده ها است را بدست می‌آوریم.

2. داده های هر کلاس را جداسازی کرده و میانگین و واریانس هر کدام را بدست می‌آوریم.

3. با فرض اینکه توزیع این کلاس ها به صورت نرمال است احتمال $p(x_j|C_j)$ را بدست می‌آوریم.

4. حال با پیدا کردن ماکسیمم رابطه $p\left(c_j\middle|X_{test}\right) \propto p_{prior} * p(X_{test}|c_j)$ به ازای کلاس های مختلف بدست می آوریم و این داده تست را به آن کلاس اختصاص می‌دهیم.

.2

```
accuracy: 0.333
```