دانشکده مهندسی کامپیوتر
دانشگاه صنعتی شریف

استاد درس: دکتر حمیدرضا ربیعی

بهار ۱۴۰۰

# تمرین در خانه چهارم
## درس یادگیری ماشین آماری


**نام و نام خانوادگی: امیر پورمند**

**شماره دانشجویی: ۹۹۲۱۰۲۵۹**

آدرس ایمیل : **pourmand1376@gmail.com**

# ۱ سوال ۱

میدانیم جواب مسئله $K^T C^{-1} y$ است.
خب ابتدا ماتریس کوورایانس ورودی ها را باید بدست بیاوریم.

$$K_{x,x'} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{۱}$$

و ماتریس کوورایانس خروجی ها میشود:

$$Cov(y_i, y_i') = \begin{bmatrix} 1+0.5^2 & 0 & 0 & 0 \\ 0 & 1+0.5^2 & 0 & 0 \\ 0 & 0 & 1+0.5^2 & 0 \\ 0 & 0 & 0 & 1+0.5^2 \end{bmatrix} \tag{۲}$$

و معکوس کوواریانس معادل زیر خواهد شد:

$$C^{-1} = \begin{bmatrix} \frac{1}{1.25} & 0 & 0 & 0 \\ 0 & \frac{1}{1.25} & 0 & 0 \\ 0 & 0 & \frac{1}{1.25} & 0 \\ 0 & 0 & 0 & \frac{1}{1.25} \end{bmatrix} = \begin{bmatrix} 0.8 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 \\ 0 & 0 & 0.8 & 0 \\ 0 & 0 & 0 & 0.8 \end{bmatrix} \tag{۳}$$

حال ماتریس کرنل ورودی را بدست می آوریم:

$$K(x, x^*) = \begin{bmatrix} 1-0.7 \\ 0 \\ 1-0.4 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.3 \\ 0 \\ 0.6 \\ 0 \end{bmatrix}$$

و نهایتا پیش بینی ما به شکل زیر خواهد بود!

$$\begin{aligned} K^T C^{-1} y &= \begin{bmatrix} 0.3 & 0 & 0.5 & 0 \end{bmatrix} \begin{bmatrix} 0.8 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 \\ 0 & 0 & 0.8 & 0 \\ 0 & 0 & 0 & 0.8 \end{bmatrix} \begin{bmatrix} 2.0 \\ 3.3 \\ 3.0 \\ 2.7 \end{bmatrix} \\ &= \begin{bmatrix} 0.24 & 0 & 0.48 & 0 \end{bmatrix} \begin{bmatrix} 2.0 \\ 3.3 \\ 3.0 \\ 2.7 \end{bmatrix} \\ &= 2 * 0.24 + 3 * 0.48 = 1.92 \end{aligned} \tag{۴}$$

## ۲ سوال ۲

خب اول سعی کنیم توزیع تجمعی $S_n$ را بدست آوریم. یعنی تابع توزیع زمان های رسیدن را بدست آوریم. ابتدا معادلات زیر را جهت فهم بهتر مسأله در نظر بگیریم. البته می دانیم $T_i$ زمان بین اتفاق $i-1$ و $i$ ام است که به آن interarrival گویند.

$$\begin{aligned}
&S_1 = T_1 \\
&S_2 = T_1 + T_2 \\
&S_3 = T_1 + T_2 + T_3 \\
&\dots \\
&S_n = \sum_{i=1}^{n} T_i
\end{aligned} \tag{۵}$$

خب داریم

$$\begin{aligned}
F_{S_n}(s) &= P(S_n \leq s) \\
&= P(N(s) \geq n) \\
&= 1 - P(N(s) < n) \\
&= 1 - \sum_{i=0}^{n-1} P(N(s) = i) \\
&= 1 - \sum_{i=0}^{n-1} \frac{e^{-s\lambda}(s\lambda)^i}{i!} \\
&= 1 - e^{-s\lambda} \sum_{i=0}^{n-1} \frac{(s\lambda)^i}{i!}
\end{aligned} \tag{۶}$$

حال برای بدست آوردن توزیع مورد نظر از این عبارت نسبت به s مشتق میگیریم

$$
\begin{aligned}
f_{S_n}(s) &= \frac{\partial}{\partial s} F_{S_n}(s) \\
&= -(-\lambda) e^{-s\lambda} \sum_{i=0}^{n-1} \frac{(s\lambda)^i}{i!} + (-e^{-s\lambda}) \sum_{i=1}^{n-1} \frac{i s^{i-1} \lambda^i}{i!} \\
&= \lambda e^{-s\lambda} - e^{-s\lambda} \sum_{i=1}^{n-1} \left( \frac{-\lambda \lambda^i s^i + i s^{i-1} \lambda^i}{i!} \right) \\
&= \lambda e^{-s\lambda} - \lambda e^{-s\lambda} \sum_{i=1}^{n-1} \left[ \frac{s^{i-1} \lambda^{i-1}}{(i-1)!} - \frac{\lambda^i s^i}{i!} \right] \\
&= \lambda e^{-s\lambda} + \lambda e^{-s\lambda} \left( \frac{(s\lambda)^{n-1}}{(n-1)!} - 1 \right) \\
&= \frac{s^{n-1} \lambda^n e^{-s\lambda}}{(n-1)!} \\
&= \frac{s^{n-1} \lambda^n e^{-s\lambda}}{\Gamma(n)}
\end{aligned} \tag{۷}
$$

پس فهمیدیم متغیر $S_n$ از توزیع گاما پیروی میکند که به شکل زیر میتوان آن را نوشت.

$$
S_n \sim Gamma(s, \lambda) \tag{۸}
$$

حال که توزیع متغیر $S_n$ بدست اومد می توان توزیع $S_n|N(t)$ را بدست آورد و این توزیع مهم است زیرا حد بالای سیگمای ما $N(t)$ است. البته اگر حدبالای سیگما یک عدد ثابت بود کار خیلی راحت بود و میتوانستیم به سادگی توزیع $S_n^2$ را بدست آورده و باهم جمع کنیم.

خب پس برویم سراغ بدست آوردن توزیع شرطی: (البته با توجه به فرض سوال مشخص است که همیشه $n \leq i$)

$$
\begin{aligned}
f_{S_i, N(t)}(s, n) &= P(S_i = s | N(t) = n) \\
&= \frac{P(N(t) = n | S_i = s) P(S_i = s)}{P(N(t) = n)} \\
&= \frac{P(N(t) - N(s) = n - i) P(S_i = s)}{P(N(t) = n)} \\
&= \frac{e^{-\lambda(t-s)} (\lambda(t-s))^{n-i}}{(n-i)!} \frac{s^{i-1} \lambda^i e^{-s\lambda}}{\Gamma(i)} \frac{n!}{e^{-\lambda t} (\lambda t)^n} \\
&= \frac{(t-s)^{n-i} s^{i-1} n!}{(n-i)! \Gamma(i) t^n} \\
&= \frac{\Gamma(n+1)}{\Gamma(n-i+1) \Gamma(i)} \frac{(t-s)^{n-i} s^{i-1}}{t^n}
\end{aligned} \tag{۹}
$$

خیلی هم خوب. با قدری محاسبات ریاضی به عبارت رو به رو رسیدیم. حالا اگر بتوانیم از دست t مخرج خلاص شویم یک توزیع تر و تمیزتر معادل توزیع بتا داریم که امیدریاضی توان دو آن بهتر محاسبه میشود. پس یک تبدیل انجام می دهیم.

ابتدا داریم $Y = S_i/t$ پس $S_i = Yt$ و البته مشتق جزئی آن میشود $\frac{\partial}{\partial Y} S_i = t$
پس تابع توزیع آن به شکل زیر خواهد بود.

$$\begin{aligned} f_{Y|N(t)}(y,n) &= f_{S_i|N(t)}(yt,n)|t| \\ &= \frac{\Gamma(n+1)}{\Gamma(n-i+1)\Gamma(i)} \frac{(t-yt)^{n-i}(yt)^{i-1}t}{t^n} \\ &= \frac{\Gamma(n+1)}{\Gamma(n-i+1)\Gamma(i)} (1-y)^{n-i} y^{i-1} \end{aligned} \tag{۱۰}$$

که البته میدانیم $0 < y < 1$ است. پس یک توزیع بتا به شکل $Beta(n-i+1, i)$ داریم که میدانیم امیدریاضی توان ۲ آن به شکل زیر خواهد بود.

$$E\left(Y^2\right) = \frac{\Gamma(n+1)\Gamma(i+2)}{\Gamma(i)\Gamma(n+3)} = \frac{i(i+1)}{(n+1)(n+2)} \tag{۱۱}$$

حال کم کم به خواسته اصلی مسأله برسیم!

$$
\begin{aligned}
\mathbb{E}[t] = \mathbb{E}\left[\sum_{i=1}^{N(t)} S_n^2\right] &= \mathbb{E}_{S_n,N(t)}\left[\sum_{i=1}^{N(t)} t^2(S_n/t)^2\right] \\
&= t^2\sum_{k=1}^{N(t)}\int_0^\infty (S_n/t)^2 f_{S_n,N(t)}(i,k)di \\
&= t^2\sum_{k=1}^{N(t)}\left[\int_0^\infty (S_n/t)^2 f_{S_n|N(t)}(i|k)di\right] f_{N(t)}(k) \\
&= t^2\sum_{k=1}^{N(t)}\mathbb{E}\left[(S_n/t)^2|N(t)=n)\right] f_{N(t)}(k) \\
&= t^2\,\mathbb{E}_{N(t)=n}\left[\mathbb{E}_{S(n)}\left[S_n/t)^2|N(t)=n\right]\right] \\
&= t^2\,\mathbb{E}_{N(t)=n}\left[\sum_{i=1}^{n}\frac{i(i+1)}{(n+1)(n+2)}\right] \\
&= t^2\,\mathbb{E}_{N(t)=n}\left[\sum_{i=1}^{n}\frac{i^2+i}{(n+1)(n+2)}\right] \\
&= t^2\,\mathbb{E}_{N(t)=n}\left[\frac{1}{(n+1)(n+2)}\sum_{i=1}^{n} i^2+i\right] \\
&= t^2\,\mathbb{E}_{N(t)=n}\left[\frac{1}{(n+1)(n+2)}(\frac{n(n+1)(2n+1)}{6}+\frac{n(n+1)}{2})\right] \\
&= t^2\,\mathbb{E}_{N(t)=n}\left[\frac{n(2n+1)}{6(n+2)}+\frac{n}{2(n+2)}\right] \\
&= t^2\,\mathbb{E}_{N(t)=n}\left[\frac{2n^2+n+3n}{6n+2)}\right] \\
&= t^2\,\mathbb{E}_{N(t)=n}\left[\frac{(n+2)n}{3(n+2)}\right] = t^2\,\mathbb{E}_{N(t)=n}\left[\frac{n}{3}\right] = \frac{\lambda t^3}{3}
\end{aligned}
\tag{۱۲}
$$

## ۳ سوال ۳

در کوییز اثبات کردم که امیدریاضی زمان رسیدن به مقصد برابر مقدار زیر خواهد بود.

$$[W - (R + \frac{1}{\lambda})]e^{-s\lambda} + (\frac{1}{\lambda} + R) \quad (۱۳)$$

و گفتیم اگر $w > R + \frac{1}{\lambda}$ باشد پس کل پرانتز اول مثبت خواهد بود و بهتر است برای مینیمم کردن مقدار $e^{-s\lambda}$ مقدار $s$ را به سمت بی نهایت ببریم. البته توجه داریم که بقیه ترم ها ثابت فرض شده اند و تاثیری ندارند.

اگر هم مقدار پرانتز منفی است یعنی $w < R + \frac{1}{\lambda}$ است بهتر است برای مینیمم کردن مقدار s را صفر بگذاریم.

حال توضیح شهودی چیست؟ مشخص است که اگر $W < R + \frac{1}{\lambda}$ باشد یعنی زمان پیاده رفتن تا رسیدن به مقصد از زمانی که طول میکشد تا منتظر اتوبوس باشیم و سپس به سمت مقصد حرکت کنیم کمتر است که هرگز اصلا منتظر اتوبوس نمی مانیم و پیاده راهی مقصد خواهیم شد یعنی s همان صفر است.

اگر هم زمان پیاده رفتن بیشتر زمان منتظر ماندن و سپس با اتوبوس رفتن است یعنی $W > R + \frac{1}{\lambda}$ در لحظه اول طرف به نفعش است که صبر کند. با توجه به خاصیت بدون حافظه بودن توزیع نمایی در هر لحظه ای که فکر کند به نفع او هست که صبر کند تا اتوبوس بیاید پس باید تا بی نهایت صبر کند و حالت دیگری نیز نمی ماند که بررسی کنیم!

یعنی مسئله به شدت ساده است: اگر پیاده زودتر میرسی خب پیاده برو وگرنه صبر کن تا اتوبوس بیاد!