درس نظریه یادگیری ماشین

تمرین سری ۱

زهرا [illegible] ۹۸۱۰۵۱۳۸

سوال ۱-۱)

ابتدا عبارت معادل x^TAx را می‌یابیم.

c_i ها همان ستون های A هستند

$$\underset{1\times1}{(x^TAx)} = (x^T)\begin{pmatrix} c_1 & \cdots & c_n \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$= \begin{pmatrix} \langle x, c_1\rangle & \langle x, c_2\rangle & \cdots & \langle x, c_n\rangle \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$= \sum_{i=1}^{n} \langle x, c_i\rangle x_i$$

$$= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} x_j c_{ji}\right) x_i$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} x_j a_{ji} x_i$$

حال برای $\frac{\partial x^TAx}{\partial x}$ داریم:

$$\frac{\partial x^TAx}{\partial x} = \left(\frac{\partial x^TAx}{\partial x_1} \cdots \frac{\partial x^TAx}{\partial x_n}\right)_{1\times n}$$

$$\forall k=1,\dots,n\ ; \quad \frac{\partial x^TAx}{\partial x_k} = \frac{\partial}{\partial x_k}\left(\sum_{i,j=1}^{n} x_i x_j a_{ji}\right)$$

سایر جملات که x_k ندارند

$$= \frac{\partial}{\partial x_k}\left(\sum_{j=1}^{n} x_k x_j a_{jk} + \sum_{i=1}^{n} x_i x_k a_{ki}\right) + 0$$

$$= \sum_{j=1}^{n} x_j a_{jk} + \sum_{i=1}^{n} x_k a_{ki}$$

$$= \langle x, c_k\rangle + \langle x, r_k\rangle$$

$$\Rightarrow \frac{\partial x^TAx}{\partial x} = \left(\langle x, c_1\rangle + \langle x, r_1\rangle, \cdots, \langle x, c_n\rangle + \langle x, r_n\rangle\right)$$

$$= \left(\langle x, c_1\rangle \cdots \langle x, c_n\rangle\right) + \left(\langle x, r_1\rangle \cdots \langle x, r_n\rangle\right)$$

$$= x^TA + x^TA^T = x^T(A + A^T)$$

$$\Rightarrow \quad \frac{\partial x^T A x}{\partial x} = x^T (A + A^T)$$

اگر A متقارن $\Rightarrow A = A^T \Rightarrow \dfrac{\partial x^T A x}{\partial x} = x^T (A + A) = 2 x^T A$

۱.۲) می‌خواهیم نشان دهیم اگر $\lambda_1, \dots, \lambda_n$ مقادیر ویژه ماتریس A باشند در این صورت

$$\sum_{i=1}^{n} \lambda_i = trace(A) \quad \text{و} \quad \prod_{i=1}^{n} \lambda_i = det(A)$$

چندجمله‌ای مشخصه ماتریس A به صورت: $P_A(\lambda) = |\lambda I_n - A|$ می‌باشد و از درجه n

پس به صورت $P_A(\lambda) = \lambda^n + \alpha_{n-1}\lambda^{n-1} + \dots + \alpha_1 \lambda + \alpha_0$

و با توجه به اینکه می‌دانیم مقادیر ویژه ماتریس A ریشه‌های غیر صفر چندجمله‌ای مشخصه هستند

پس چندجمله‌ای مشخصه را می‌توانیم فرم زیر هم نوشت.

$$P_A(\lambda) = (\lambda - \lambda_1) \dots (\lambda - \lambda_n)$$

حال برای ضریب $\alpha_0 = P_A(0)$ در چندجمله‌ای مشخصه داریم:

$$P_A(0) = \alpha_0 = (0 - \lambda_1) \dots (0 - \lambda_n) = (-1)^n \prod_{i=1}^{n} \lambda_i$$

$$P_A(0) = |\overbrace{0 I_n}^{0} - A| = (-1)^n |A|$$

$$\Rightarrow (-1)^n \prod_{i=1}^{n} \lambda_i = (-1)^n |A| \Rightarrow \prod_{i=1}^{n} \lambda_i = det(A) \quad ☑$$

حال برای اثبات قسمت دوم داریم.

ابتدا با توجه به چندجمله‌ای مشخصه به فرم $P_A(\lambda) = (\lambda - \lambda_1) \dots (\lambda - \lambda_n)$ می‌دانیم ضریب

α_{n-1} به صورت زیر قابل محاسبه است. (که از $n-1$ تا پرانتز λ و از یکی اسکالر λ_i انتخاب)

$$\Rightarrow \alpha_{n-1} = -(\lambda_1 + \dots + \lambda_n) \Rightarrow \sum_{i=1}^{n} \lambda_i = -\alpha_{n-1} \quad (1)$$

حال از طریق $|\lambda I_n - A|$ ضریب α_{n-1} را محاسبه می‌کنیم.

$$det(\lambda I_n - A) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \dots & -a_{1n} \\ \vdots & & & \\ -a_{n1} & & & \lambda - a_{nn} \end{vmatrix}$$

دترمینان را به این گونه حساب می کنیم که حاصل ضرب درایه های $a_{1i_1}, \ldots, a_{ni_n}$

که $i_1, \ldots, i_n$ یک جایگشت از اعداد $1, \ldots, n$ هستند را حساب می کنیم و

سپس این $n!$ عدد مختلف را باهم جمع می کنیم، البته با توجه به فرد یا زوج بودن

جایگشت [(1-) برای جایگشت های فرد] نیز علامت در نظر می گیریم.

از این جایگشت ها یکی به شکل $(\lambda - a_{11}) \ldots (\lambda - a_{nn})$ و بقیه جایگشت ها ماکسیمم شامل $n-2$ تا از عناصر روی قطر اصلی هستند پس ماکسیمم حداکثر λ^{n-2} می سازند. پس برای ضریب جمله λ^{n-1} (a_{n-1}) در دترمینان فقط در بخش $(\lambda - a_{11}) \ldots (\lambda - a_{nn}) = h(\lambda)$ باید بدست آوریم.

در نتیجه داریم:

$$a_{n-1} = h(\lambda) \text{ در } \lambda^{n-1} \text{ ضریب } = (-a_{11} - a_{22} - \ldots - a_{nn}) = -\sum_{i=1}^{n} a_{ii} \quad (2)$$

$$(1),(2) \Rightarrow a_{n-1} = -\sum_{i=1}^{n} \lambda_i \quad \text{و} \quad a_{n-1} = -\sum_{i=1}^{n} a_{ii} = -\operatorname{trace}(A)$$

$$\Rightarrow \sum_{i=1}^{n} \lambda_i = \operatorname{trace}(A) \quad \boxtimes$$

سوال ۱.۳)

$$f_{\mu,X,Y}(t,x,y) = f_{\mu}(t)\, f_{X|\mu}(x|t) \,.\, f_{Y|\mu}(y|t)$$

$\Rightarrow$ در واقع: با توجه به مستقل بودن X, Y

$$f_{\mu,X,Y}(t,x,y) = f_{XY|\mu}(x,y|t) \,.\, f_{\mu}(t)$$

$$f_{\mu XY}(t,x,y) = f_{X|\mu}(x|t)\, f_{Y|\mu}(y|t)\, f_{\mu}(t) \quad \nabla =$$ استقلال X,Y

با توجه به $\mu \sim Uni(0,1)$ $\Rightarrow$

$$f_{\mu}(t) = \begin{cases} 1 & t \in [0,1] \\ 0 & \# \end{cases}$$

برای توزیع های $f_{X|\mu}(x|t)$, $f_{Y|\mu}(y|t)$ در واقع با شرط اینکه $\mu = t$ باشد X, Y متغیرهای نرمال با میانگین t و واریانس 1 هستند.
پس برای توزیع شان داریم:

$$f_{X|\mu}(x|t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2}}$$

$$f_{Y|\mu}(y|t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-t)^2}{2}}$$

$$\Rightarrow f_{\mu X,Y}(t,x,y) = \begin{cases} \frac{1}{2\pi} e^{-\frac{(x-t)^2+(y-t)^2}{2}} & t \in [0,1] \\ 0 & \# \end{cases}$$

با فرض داشتن $X=x_0 , Y=y_0$ به دنبال Q هستیم که

$$Z = f_{X,Y|\mu}(x_0, y_0 | Q)\, f_{\mu}(Q)$$

ماکسیمم شود

$$\Rightarrow Z = f_{X,Y,\mu}(x_0, y_0, Q)$$

$$Z = f_{X,Y,\mu}(x_0, y_0, Q) = \begin{cases} \dfrac{e^{-\frac{(x_0-Q)^2+(y_0-Q)^2}{2}}}{2\pi} & Q\in[0,1] \\ 0 & \text{otherwise} \end{cases}$$

به دنبال Q که مقدار Z را max کند می‌گردیم [که در $Q\in[0,1]$ اتفاق افتد چون مقدار $+$ دارد در مقابل صفر]

پس از Z مشتق می‌گیریم، داریم:

$$\frac{dZ}{dQ} = \frac{(-2Q+x_0+y_0)\, e^{-\frac{(x_0-Q)^2+(y_0-Q)^2}{2}}}{2\pi} = 0$$

$$\Rightarrow -2Q+x_0+y_0=0 \Rightarrow Q=\frac{x_0+y_0}{2}$$

پس max در نقطه‌ی $(Q=\frac{x_0+y_0}{2})$ رخ می‌دهد

پس تخمین به صورت MAP، $\frac{x_0+y_0}{2}$ می‌باشد.

سوال ۱.۴)

$$A=\begin{bmatrix}5 & -4 & 2\\ -1 & 2 & 3\\ -2 & 1 & 0\end{bmatrix}$$

ماکسیمم در ستون اول

$$\|A\|_1=\max_{1\le j\le n}\sum_{i=1}^{n}|a_{ij}|=\{|5|+|-1|+|-2|\}=8$$

ماکسیمم در سطر اول

$$\|A\|_\infty=\max_{1\le i\le n}\sum_{j=1}^{n}|a_{ij}|=\{|5|+|-4|+|2|\}=11$$

$$\|A\|_2=\sqrt{\lambda_{max}(A^TA)}$$

$$A^TA=\begin{bmatrix}5 & -1 & -2\\ -4 & 2 & 1\\ 2 & 3 & 0\end{bmatrix}\begin{bmatrix}5 & -4 & 2\\ -1 & 2 & 3\\ -2 & 1 & 0\end{bmatrix}=\begin{bmatrix}30 & -24 & 7\\ -24 & 21 & -2\\ 7 & -2 & 13\end{bmatrix}$$

از ریشه های معادله مشخصه A، λ_{max} را می یابیم

$|A-\lambda I|=0 \Rightarrow$ ماکسیمم ریشه ها $=\lambda_{max}$

$$\Rightarrow -\lambda^3+64\lambda^2-664\lambda+225=0 \Rightarrow \lambda_{max}\simeq 51.058$$

$$\Rightarrow \|A\|_2=\sqrt{51.058}=7.147$$

(۲.۱۰.۱۳)

$$f(x) = \begin{cases} 2x-2 & 1<x<2 \\ 0 & \# \end{cases}$$

$$E(\underbrace{x^3+2x-7}_{g(x)}) = \int_{-\infty}^{+\infty} g(x)\, f_X(x)\, dx$$

$$= \int_{-\infty}^{+\infty} (x^3+2x-7)\, f_X(x)\, dx$$

$$= \int_1^2 (x^3+2x-7)(2x-2)\, dx + \cancelto{0}{\int_{x\in\#} (x^3+2x-7)\times 0\, dx}$$

$$= \int_1^2 (2x^4+4x^2+14-18x-2x^3)\, dx$$

$$= \left(\frac{2x^5}{5} + \frac{4x^3}{3} - 9x^2 - \frac{x^4}{2} + 14x\right)\Bigg|_1^2$$

$$= \frac{37}{30}$$

سوال ۲.۲)

ابتدا CDF آن را حساب کرده سپس با مشتق گیری PDF را محاسبه می‌کنیم.

$Y_1 = \max[X_1, \dots, X_n]$

$$F_{Y_1}(y) = \mathbb{P}[Y \leq y] = \mathbb{P}[X_1 \leq y, \dots, X_n \leq y]$$

$$= \mathbb{P}[X_1 \leq y] \dots \mathbb{P}[X_n \leq y] \quad \text{استقلال } X_i \text{ ها}$$

$$= \prod_{i=1}^{n} F_{X_i}(y)$$

$$= (F_X(y))^n \quad X_i \text{ ها از یک توزیع آمده‌اند iid}$$

$$\Rightarrow f_{Y_1}(y) = (F_{Y_1}(y))' = (F_X^n(y))' = n f_X(y) F_X(y)^{n-1}$$

$Y_2 = \min[X_1, \dots, X_n]$

$$F_{Y_2}(y) = \mathbb{P}[Y \leq y] = 1 - \mathbb{P}[Y > y]$$

$$= 1 - \mathbb{P}[X_1 > y, \dots, X_n > y]$$

$$= 1 - \prod_{i=1}^{n} \mathbb{P}[X_i > y] \quad \text{استقلال } X_i \text{ ها}$$

$$= 1 - \prod_{i=1}^{n} (1 - \mathbb{P}[X_i \leq y])$$

$$= 1 - \prod_{i=1}^{n} (1 - F_X(y))$$

$$= 1 - (1 - F_X(y))^n \quad \text{iid: } X_i \text{ ها از یک توزیع}$$

$$\Rightarrow F_{Y_2}(y) = 1 - (1 - F_X(y))^n$$

$$\Rightarrow f_{Y_2}(y) = (F_{Y_2}(y))' = (1 - (1 - F_X(y))^n)' = n f_X(y) \cdot (1 - F_X(y))^{n-1}$$

سوال ۲.۳)

نشان بدهیم که مجموعه مقادیر ویژه دو ماتریس M و $P^{-1}MP$ یکسان هستند.
به این صورت که:

$\forall \lambda \in$ مقادیر ویژه $M \overset{?}{\Rightarrow} \lambda \in$ مقادیر ویژه $P^{-1}MP$

و $\forall \lambda \in$ مقادیر ویژه $P^{-1}MP \overset{?}{\Rightarrow} \lambda \in$ مقادیر ویژه M

ابتدا حکم دوم را ثابت می‌کنیم، یعنی فرض می‌کنیم λ مقدار ویژه ماتریس $P^{-1}MP$ می‌باشد
نشان می‌دهیم مقدار ویژه‌ای برای ماتریس M نیز هست.

فرض: λ مقدار ویژه ماتریس $P^{-1}MP \Rightarrow \exists X \neq 0;\ P^{-1}MPX = \lambda X \xRightarrow{PX} \underbrace{PP^{-1}}_{I}M\overbrace{PX}^{Y} = \lambda \overbrace{PX}^{Y \neq 0}$

$\xRightarrow[(PX=Y)]{\text{بنامیم}} MY = \lambda Y \Rightarrow$ λ مقدار ویژه ماتریس M است.

حکم: اول: به طور مشابه ابتدا فرض می‌کنیم λ مقدار ویژه ماتریس M می‌باشد سپس نشان می‌دهیم
λ مقدار ویژه‌ای برای ماتریس $P^{-1}MP$ نیز می‌باشد.

فرض: λ مقدار ویژه ماتریس $M \Rightarrow \exists X \neq 0;\ MX = \lambda X$ ①
با توجه به full rank بودن ماتریس P می‌دانیم برای معادله $PY = X$ جواب
یکتا $Y \neq 0$ وجود دارد. $\exists Y \neq 0;\ PY = X$ به صورت یکتا.

در نتیجه: به جای X، PY قرار می‌دهیم و داریم:

$① \Rightarrow MPY = \lambda PY \xRightarrow{P^{-1}\times} P^{-1}MPY = \lambda Y$

پس λ مقدار ویژه‌ای برای ماتریس $P^{-1}MP$ می‌باشد.

پس به طور کامل حکم ثابت می‌شود و مجموعه مقادیر ویژه دو ماتریس یکی است.

P4PCO