Introduction to Machine Learning
Fall 2019
University of Science and Technology of China

Lecturer: Jie Wang
Posted: Dec. 9, 2019
Name: Bowen Zhang

Homework 7
Due: Dec. 23, 2019
ID: PB17000215

Notice, to get the full credits, please present your solutions step by step.

Exercise 1: Properties of expectation and variance 10pts

Let X, Y, and Z be random variables. Show that the following results hold.

1. (5pts) The tower property holds, i.e.,

$$\mathrm{E}[X|Y] = \mathrm{E}[\mathrm{E}[X|Y, Z]|Y].$$

2. (5pts) The variance decomposition formula holds, i.e.,

$$\mathrm{Var}[X] = \mathrm{E}[\mathrm{Var}[X|Y]] + \mathrm{Var}[\mathrm{E}[X|Y]].$$

Hint: if you do not know measure theory well, you can assume that X, Y, and Z are continuous random variables.

1、令 $g(Y,Z) = E[X|Y=y, Z=z]$

$$\begin{aligned}
E[E[X|Y,Z]|Y] &= E[g(Y,Z)|Y=y] \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(y',z) f(y',z|Y=y)\,dy'dz \\
&= \int_{-\infty}^{+\infty} g(y,z) f(z|Y=y)\,dz \\
&= \int_{-\infty}^{+\infty} E[X|Y=y, Z=z] f(z|Y=y)\,dz \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x f(x|Y=y, Z=z) f(z|Y=y)\,dx\,dz \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x f(x,z|Y=y)\,dx\,dz \\
&= \int_{-\infty}^{+\infty} x f(x|Y=y)\,dx \\
&= E[X|Y]
\end{aligned}$$

2.
$$\begin{aligned}
Var[X] &= EX^2 - (EX)^2 \\
&= E[E[X^2|Y]] - (E[E[X|Y]])^2 \\
&= E[var[X|Y] + (E[X|Y])^2] - (E[E[X|Y]])^2 \\
&= E[var[X|Y]] + E[(E[X|Y])^2] - (E[E[X|Y]])^2 \\
&= E[var[X|Y]] + var[E[X|Y]]
\end{aligned}$$

$$\|M\|_\infty = \max_{\|\mathbf{x}\|_\infty \le 1} \|M\mathbf{x}\|_\infty.$$

i. Show that $\|M\|_\infty = \max_i \sum_{j=1}^n |m_{i,j}|$.

ii. Show that $\|cM\|_\infty = |c|\|M\|_\infty$ for any $c \in \mathbb{R}$.

iii. Show that $\|AB\|_\infty \le \|A\|_\infty \|B\|_\infty$ holds for any $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$.

(b) Show that the sequence $I, \gamma T, \sum_{i=0}^2 (\gamma T)^2, \ldots, \sum_{i=0}^n (\gamma T)^n, \ldots$ is Cauchy. (Hint: a matrix sequence $\{A_p\}$ is Cauchy in $(\mathbb{R}^{m\times n}, \|\cdot\|_\infty)$, if given any $\epsilon > 0$, there is an integer $N \ge 1$ such that $\|A_p - A_q\|_\infty < \epsilon$ whenever $p, q \ge N$.)

(c) Combining the result in the last part and the fact that $\mathbb{R}^{n\times n}$ is complete, we can conclude that $\sum_{i=0}^\infty (\gamma T)^i$ converges to a matrix which we denote by L. Show that $(I - \gamma T)^{-1} = L$. (Hint: you need to show that $(I - \gamma T)L = \lim_{n\to\infty} \sum_{i=0}^n (I - \gamma T)(\gamma T)^i$.)

(1) 取 $x=(1,1,\cdots 1)^T \in R^{n\times 1}$

$T x = (\sum_j t_{1j}, \sum_j t_{2j} \cdots \sum_j t_{nj})^T$

由其性质知 $= (1, 1, \cdots 1)^T = x$

故有 $\lambda=1$ 的特征值.

(2) 若T是右随机矩阵则 T^k 也为右随机矩阵 $(k \ge 1)$

验证: ~~若~~取 $x=(1,1\ldots)^T \in R^{n\times 1}$

$T^k x = T^{k-1} T x = T^{k-1} x = \cdots = T x = x$

各行之和为1, 且 T^k 的各元素均非负, 故 T^k 也为右随机矩阵.

假设存在 $|\lambda|>1$ 的特征值, 有特征向量 x'

则 $T x' = \lambda x'$

$\vdots$

$T^k x' = \lambda^k x'$

$k\to\infty$ 时, 说明 T^k 至少有一个元素 >1, 但这与 T^k 各元素非负, 各行元素之和为1, 矛盾, 故 $|\lambda| \le 1$.

(3) 假设~~其~~ $I-\gamma T$ 有0特征值.

则存在 x 使得

$(I-\gamma T)x = 0$

$\gamma T x = I x$

$T x = \frac{1}{\gamma} x$

$|\frac{1}{\gamma}|>1$, 这说明 T 有大于1的特征值, 与(2)矛盾, 故 $I-\gamma T$ 特征值均非零, 故可逆.

4.

(a)

i. $\|M\|_\infty = \max\limits_{i, |x_j|\le 1} \left|\sum\limits_{j=1}^{n} m_{ij} x_j\right| \le \max\limits_{i} \sum\limits_{j=1}^{n} |m_{ij}|$

取 $x_j = \mathrm{sgn}(m_{ij})$ $\quad 1\le i\le m, 1\le j\le n,\ i,j\in N$

则可以取到等于号

故 $\|M\|_\infty = \max\limits_{i} \sum\limits_{j=1}^{n} |m_{ij}|$

ii. $\|cM\|_\infty = \max\limits_{i} \sum\limits_{j=1}^{n} |c m_{ij}| = |c| \max\limits_{i} \sum\limits_{j=1}^{n} |m_{ij}| = |c|\,\|M\|_\infty$

iii. $(AB)_{ij} = \sum\limits_{k=1}^{n} a_{ik} b_{kj}$

$$
\begin{aligned}
\|AB\|_\infty &= \max_{i} \sum_{j=1}^{p} \left|\sum_{k=1}^{n} a_{ik} b_{kj}\right| \\
&\le \max_{i} \sum_{j=1}^{p} \sum_{k=1}^{n} |a_{ik}|\,|b_{kj}| \\
&= \max_{i} \sum_{k=1}^{n} |a_{ik}| \sum_{j=1}^{p} |b_{kj}| \\
&\le \max_{i} \sum_{k=1}^{n} |a_{ik}| \max_{t} \sum_{j=1}^{p} |b_{tj}| \\
&= \|A\|_\infty \|B\|_\infty
\end{aligned}
$$

(b) $A_p - A_q = \sum\limits_{i=q+1}^{p} (\gamma T)^i$

由于其元素均非负，取 $x = (1,1,\cdots 1)^T \in R^{n\times 1}$

$$
\begin{aligned}
(A_p - A_q)x &= \sum_{i=q+1}^{p} \gamma^i T^i x \\
&= \sum_{i=q+1}^{p} \gamma^i x \quad (\text{利用 } T^i \text{ 也为随机矩阵及其性质})
\end{aligned}
$$

$\|A_p - A_q\|_\infty = \max\limits_{\|x\|_\infty \le 1} \|(A_p - A_q)x\|_\infty$

参考 (a).i 可得

$$
\begin{aligned}
&= \max_{x=(1,1,\cdots 1)^T} \left\|\sum_{i=q+1}^{p} \gamma^i x\right\|_\infty \\
&= \sum_{i=q+1}^{p} \gamma^i \\
&= \frac{\gamma^{q+1}(1-\gamma^{p-q})}{1-\gamma} < \frac{\gamma^{q+1}}{1-\gamma} < \varepsilon
\end{aligned}
$$

得 $q > \dfrac{\ln \varepsilon(1-\gamma)}{\ln \gamma} - 1$

故取 $N = \left\lceil \dfrac{\ln \varepsilon(1-\gamma)}{\ln \gamma} - 1 \right\rceil$ 那么 $p, q \ge N$ 时 $\|A_p - A_q\|_\infty < \varepsilon$

(c)

考察

$$\sum_{i=0}^{n}(I-\gamma T)(\gamma T)^{i}$$

$$=\sum_{i=0}^{n}(\gamma T)^{i}-(\gamma T)^{i+1}$$

$$= I-(\gamma T)+(\gamma T)-(\gamma T)^{2}+\cdots-(\gamma T)^{n+1}$$

$$= I-(\gamma T)^{n+1}$$

因 T^{n+1} 各元素大小有限(右随机矩阵性质)

$0<\gamma<1$，故 $(\gamma T)^{n+1}\to 0$ 当 $n\to\infty$ 时

$$\therefore \lim_{n\to\infty}\sum_{i=0}^{n}(I-\gamma T)(\gamma T)^{i} = I$$

$\because$ $I-\gamma T$ 可逆

$$\therefore \lim_{n\to\infty}\sum_{i=0}^{n}(\gamma T)^{i}=(I-\gamma T)^{-1}$$

3.

1.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}
S_1	0.15	0.05									
S_2	0.8	0.9	0.05								
S_3		0.05	0.15								
S_4	0.05			0.2		0.05					
S_5			0.8		0.2						
S_6				0.8		0.1			0.8		
S_7						0.8	0.15	0.05			
S_8					0.8		0.8	0.15			
S_9						0.05			0.15		
S_{10}							0.05		0.05	1	
S_{11}								0.8			1

2.

M^T 为状态转移矩阵.

(a) 开始的状态向量为 $x=(\frac{1}{11},\frac{1}{11}\cdots\frac{1}{11})$

$$P(S_1)=xM^T=(0.0182,0.159,0.0182,0.0273,0.091,0.155,0.091,0.159,0.0182,0.1,0.164)$$

$$P(S_2)=xM^TM^T=(0.0107,0.159,0.011,0.041,0.0327,0.052,0.145,0.169,0.0105,0.106,0.291)$$

(b) 考察 M^T 的特征值分解，用 numpy 分解得

$$M^T=U\Sigma U^{-1}$$

$$\Sigma=diag(0.953,0.097,0.15,-0.149,0.174,0.424,-0.05,0.35,0.2,1,1)$$

可见除了最后两个特征值 $=1$ 之外，前面的特征值均 <1.

$$P(S_t)=x(M^T)^t$$
$$=x(U\Sigma U^{-1})^t$$
$$=xU\Sigma^tU^{-1}$$

当 $t\to\infty$ 时，$\Sigma^t\to diag(0,0,0\cdots,1,1)$

可见 $P(S_t=S_i)=0 \quad i=1,2,\cdots 9$

(c) 同上，$\lim\limits_{t\to\infty}P(S_t=S_{10})=\lim\limits_{t\to\infty}(xU\Sigma^tU^{-1})_{10}=0.122$

同理 $\lim\limits_{t\to\infty}P(S_t=S_{11})=\lim\limits_{t\to\infty}(xU\Sigma^tU^{-1})_{11}=0.878.$

3、

$E(S_8) = 0.8 \times 100 = 80$

$E(S_7) = 0.05 \times (-100) = -5$

$E(S_9) = 0.05 \times (-100) = -5$

故 Bellman 方程可以写为

$$V^{\pi} = \begin{pmatrix} V^{\pi}(S_0) \\ V^{\pi}(S_1) \\ V^{\pi}(S_2) \\ V^{\pi}(S_3) \\ \vdots \\ V^{\pi}(S_{11}) \end{pmatrix} = \begin{array}{l} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ -5 \\ 80 \\ -5 \\ \vdots \\ 0 \end{pmatrix} \end{array} \begin{array}{l} \\ \\ \\ \\ \\ \leftarrow E(S_7) \\ \leftarrow E(S_8) \\ \leftarrow E(S_9). \\ \\ \\ \end{array} + 0.9 M^T \begin{pmatrix} V^{\pi}(S_0) \\ V^{\pi}(S_1) \\ V^{\pi}(S_2) \\ \vdots \\ V^{\pi}(S_{11}) \end{pmatrix}$$

4.

由(2b)可以看出，$t \to \infty$ 时 $\Sigma^t \to diag(0,0,0 \cdots 0,1,1)$

故无论初始状态如何

$x U \Sigma^t U^{-1}$ 的前 9 个元素均为 0.

$$\lim_{t \to \infty} P(S_t = S_i) = 0, \quad i = 1, \dots 9$$

4.

1.

$$V^* = E[r(s,\pi^*(s))] + \gamma \sum_{s'} P(s'|s,\pi^*(s))\, V^*(s')$$

$$Q(s,a) = E[r(s,a)] + \gamma \sum_{s'} [P(s'|s,a) \max_{a'} Q(s',a')]$$

2.

Value Iteration:

Initialize $V(s)$ to arbitrary values
while not terminate
  For $s \in S$
    For $a \in A$
      $Q(s,a) \leftarrow E[r(s,a)] + \gamma \sum_{s'} P(s'|s,a)V(s')$
    $V(s) \leftarrow \max_a Q(s,a)$

Policy Iteration:

Initialize $\pi \leftarrow \pi_2$, $\pi' \nLeftarrow \pi_2$
while(π != π')
  $V \leftarrow (I - \gamma T^{\pi})^{-1} R^{\pi}$
  $\pi' \leftarrow \pi$
  For $s \in S$
    $\pi(s) \leftarrow \arg\max_a E[r(s,a)] + \gamma \sum_{s'} P(s'|s,a)V(s')$

3. 设定 $r(S_{10}) = -N$ 其中 N 为一个足够大的正数.

在 Ex 3.4 中已经证明了 $\lim_{t\to\infty} P(S_t = S_i) = 0 \quad i = 1, 2, \cdots, 9$

只需证明在 $P(S_0 = S_{10}) = P(S_0 = S_{11}) = 0$ 下

$\lim_{t\to\infty} P(S_t = S_{10}) = 0.$

这等价于证明最佳策略下, S_7 选择向左, S_9 选择向下. 否则, $P(S_t = S_{10}) \neq 0$

首先假设最佳策略 S_7 选择的不是向左. 使用 Policy Iteration.

$$V = (I - \gamma T^\pi)^{-1} R^\pi \qquad 其中 \quad R^\pi = \begin{pmatrix} R^\pi(S_0) \\ R^\pi(S_1) \\ \vdots \\ R^\pi(S_7) \\ R^\pi(S_8) \\ \vdots \\ R^\pi(S_{11}) \end{pmatrix} \quad 其中 \; R^\pi(S_7) > -N$$

$$= \sum_{i=0}^{\infty} (\gamma T^\pi)^i R^\pi$$

$$-N < V^\pi(S_7) \leq P_{76} R^\pi(S_6) + P_{78} R^\pi(S_8) + P_{77} R^\pi(S_7)$$

$$\pi(S_7) = \arg\max_a \; E[r(s,a)] + \gamma \sum_{s'} P(s'|s,a) V(s')$$

$Q(S_7, up) = 0 + \gamma(0.8V^\pi(S_8) + 0.05(-N) + 0.15V^\pi(S_7))$

$Q(S_7, down) = 0 + \gamma(0.8V^\pi(S_6) + 0.05(-N) + 0.15V^\pi(S_7))$

$Q(S_7, right) = -N + \gamma(0.8(-N) + 0.1(V^\pi(S_7)) + 0.05(V^\pi(S_8) + V^\pi(S_6)))$

$Q(S_7, left) = 0 + \gamma(0.9V^\pi(S_7) + 0.05(V^\pi(S_6) + V^\pi(S_8)))$

~~由于 $\gamma = 0$~~

当N选择得足够大时，就有

$Q(S_7, left) > Q(S_7, up), Q(S_7, down), Q(S_7, right)$

故π会发生改变，与其中最优策略矛盾

同理若 S_7 不选择向下走也是如此，

综上 S_7 会选择向左，S_9 向下

所以 $\lim_{t\to\infty} P(S_0 = S_{10}) = 0$.

5.

1.

根据 $Q^{\pi}(s,a)$ 的定义有

$$Q^{\pi}(s,a) = E[r(s,a)] + \gamma \sum_{s'} P(s'|s,a) V^{\pi}(s')$$

而

$$V^{\pi}(s) = E[r(s,a)] + \gamma \sum_{s'} P(s'|s,a) V^{\pi}(s')$$

那么

$V^{\pi}(s')$ 可以写为 $Q^{\pi}(s',a')$

其中 $a' = \pi(s')$

故

$$Q^{\pi}(s,a) = E[r(s,a)] + \gamma \sum_{s'} P(s'|s,a) Q^{\pi}(s',a')$$

其中 $a' = \pi(s')$

2.

$\because \pi'(s) = \underset{a \in A}{\arg\max}\, Q^{\pi}(s,a)$

$\therefore \quad Q^{\pi'}(s,\pi'(s)) \geq Q^{\pi}(s,\pi(s))$

由1.可知

$$V^{\pi'}(s) = Q^{\pi'}(s,\pi'(s)) \geq Q^{\pi}(s,\pi(s)) = V^{\pi}(s)$$