Homework 1 Solution

Advanced Machine Learning 10-716

Due at **11:59pm on Tuesday, Feb 5**.
You are required to start the solution to each problem on a new page. Please make sure to assign the pages to their respective questions on Gradescope.

1 Bayesian Analysis I (Shawn Lyu)

1. Write out the posterior expected loss as

$$\int_\Theta w(\theta)(\theta - a)^2 dF^{\pi(\theta|x)}(\theta)$$

The loss, a weighted sum of convex functions, is a Take the derivative wrt a and set it to 0 we have

$$0 = \frac{d}{da}\int_\Theta w(\theta)(\theta - a)^2 dF^{\pi(\theta|x)}(\theta)$$

which can be expanded to

$$0 = \frac{d}{da}\int_\Theta \left(w(\theta)\theta^2 - 2aw(\theta)\theta + a^2 w(\theta)\right)^2 dF^{\pi(\theta|x)}(\theta) \quad = 2\mathbb{E}^{\pi(\theta|x)}[\theta w(\theta)] - 2a\mathbb{E}^{\pi(\theta|x)}[w(\theta)]$$

Solving this just shows that the Bayes rule is

$$\delta^\pi(x) = \frac{\mathbb{E}^{\pi(\theta|x)}[\theta w(\theta)]}{\mathbb{E}^{\pi(\theta|x)}[w(\theta)]}$$

2. (a) From the Berger book, section 4.4.2 result 3, we know that the Bayes rule for this kind of loss is

$$\delta^\pi(x) = \mathbb{E}^\pi[\theta]$$

The goal is then for us to find the posterior distribution $\pi(\theta|X = x)$. Since Beta distribution is the conjugate prior to binomial distribution we know that

$$\theta|X \sim Beta(x + 3, 14 - x)$$

and that

$$\delta^\pi(x) = \frac{x+3}{17}$$

(b) From the Berger book, section 4.4.2 result 5 the Bayes estimator is the median of the posterior, which tells us directly that

$$\delta^\pi(x) = \frac{3x+8}{49}$$

(c) From the Berger book, section 4.4.2 result 4 (or the result above), we know that the Bayes rule for weighted loss is

$$\delta^\pi(x) = \frac{\mathbb{E}^{\pi(\theta|x)}[\theta w(\theta)]}{\mathbb{E}^{\pi(\theta|x)}[w(\theta)]}$$

Note that since $\theta|x \sim Beta(x+3, 14-x)$, expanding the pdf yields

$$\begin{aligned}
&\mathbb{E}^{\pi(\theta|x)}[w(\theta)] \\
=&\mathbb{E}^{\pi(\theta|x)}\left[\frac{1}{\theta(1-\theta)}\right] \\
=&\int \frac{\theta^{x+2}(1-\theta)^{13-x}}{B(x+3,14-x)} \frac{1}{\theta(1-\theta)} d\theta \\
=&\int \frac{\Gamma(17)}{\Gamma(x+3)\Gamma(14-x)} \theta^{x+1}(1-\theta)^{12-x} d\theta \\
=&\frac{\Gamma(17)\Gamma(x+2)\Gamma(13-x)}{\Gamma(15)\Gamma(x+3)\Gamma(14-x)} \int \frac{\Gamma(15)}{\Gamma(x+2)\Gamma(13-x)} \theta^{x+1}(1-\theta)^{12-x} d\theta \\
=&\frac{16!}{14!(x+2)(13-x)}
\end{aligned}$$

We then find the weighted expected value of θ, which is just

$$\begin{aligned}
&\mathbb{E}^{\pi(\theta|x)}[\theta w(\theta)] \\
=&\int \frac{\Gamma(17)}{\Gamma(x+3)\Gamma(14-x)} \theta^{x+2}(1-\theta)^{13-x} \frac{1}{1-\theta} d\theta \\
=&\frac{\Gamma(17)\Gamma(x+3)\Gamma(13-x)}{\Gamma(16)\Gamma(x+3)\Gamma(14-x)} \int \frac{\Gamma(16)}{\Gamma(x+3)\Gamma(13-x)} \theta^{x+2}(1-\theta)^{12-x} d\theta \\
=&\frac{16}{(13-x)}
\end{aligned}$$

Plug these two together we have

$$\delta^\pi(x) = \frac{x+2}{15}$$

2 Bayesian Analysis II (Shawn Lyu)

1. Based on the constraint write the decision rule as

$$\delta^\pi(x) = (a_1, 1 - a_1)^T$$

we have that the loss can be written as

$$\int_\Theta \theta_1^2 a_1^2 + \theta_2^2 + \theta_2^2 a_1^2 + 1 + 2\theta_1\theta_2 a_1 - 2\theta_1\theta_2 a_1^2 - 2\theta_1 a_1 - 2\theta_2^2 a_1 - 2\theta_2 + 2\theta_2 a_1 dF^\pi(\theta)$$

Set the derivative wrt to a_1 to 0 we have

$$0 = 2\mathbb{E}[\theta_1^2]a_1 + 2\mathbb{E}[\theta_2^2]a_1 + 2\mathbb{E}[\theta_1\theta_2] - 4\mathbb{E}[\theta_1\theta_2]a_1 - 2\mathbb{E}[\theta_1] - 2\mathbb{E}[\theta_2^2] + 2\mathbb{E}[\theta_2]$$

Since the conjugate prior of Gaussian is Gaussian, we have

$$\theta|x \sim \mathcal{N}\left(\frac{x+\mu}{2}, \frac{1}{2}I_2\right)$$

We then have

$$\begin{aligned}
\mathbb{E}[\theta_1^2] &= \frac{(x_1+\mu_1)^2+2}{4} \\
\mathbb{E}[\theta_2^2] &= \frac{(x_2+\mu_2)^2+2}{4} \\
\mathbb{E}[\theta_1] &= \frac{x_1+\mu_1}{2} \\
\mathbb{E}[\theta_2] &= \frac{x_2+\mu_2}{2} \\
\mathbb{E}[\theta_1\theta_2] &= \frac{(x_1+\mu_1)(x_2+\mu_2)}{4}
\end{aligned}$$

Plug these back on top to obtain

$$\frac{(x_1+\mu_1)^2 + (x_2+\mu_2)^2 + 4 - 2(x_1+\mu_1)(x_2+\mu_2)}{2} a_1 = -\frac{(x_1+\mu_1)(x_2+\mu_2)}{2} + x_1 + \mu_1 + \frac{(x_2+\mu_2)^2}{2} + 1 - x_2 - \mu_2$$

So

$$a_1 = \frac{(x_2+\mu_2)^2 + 2 + 2(x_1+\mu_1) - 2(x_2+\mu_2) - (x_1+\mu_1)(x_2+\mu_2)}{(x_1+\mu_1)^2 + (x_2+\mu_2)^2 + 4 - 2(x_1+\mu_1)(x_2+\mu_2)}$$

and the other parameter is

$$a_2 = \frac{(x_1+\mu_1)^2 + 2 + 2(x_2+\mu_2) - 2(x_1+\mu_1) - (x_1+\mu_1)(x_2+\mu_2)}{(x_1+\mu_1)^2 + (x_2+\mu_2)^2 + 4 - 2(x_1+\mu_1)(x_2+\mu_2)}$$

Finally, our decision rule is just

$$\delta^\pi(x) = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

2. Expand the loss

$$\int_\Theta \theta^T Q\theta - 2\theta^T Qa + a^T Qa dF^\pi(\theta)$$

Find the derivative wrt a and set to 0 to obtain

$$0 = 2\mathbb{E}[\theta^T]Q - 2a^T Q$$

We then easily know that

$$\delta^\pi(x) = \mathbb{E}^\pi[\theta]$$

3 Minimax Analysis I (Karthika Nair and Yao-Hung Tsai)

1. Let δ' be another decision rule. Since δ is admissible, we can always find a point θ' depending on δ' s.t.

$$R(\theta_0, \delta) \leq R(\theta', \delta').$$

Since δ has constant risk, we have the following inequalities

$$\sup_{\theta} R(\theta, \delta) = R(\theta_0, \delta) \leq R(\theta', \delta') \leq \sup_{\theta} R(\theta, \delta').$$

(4 pts) Note that δ' can be arbitrary decision rule. Therefore, δ is minimax. (1 pts) If unique Bayes, then unique minimax.

2. (a) A = { a_1, a_2, a_3 }
where a_1 = Order of 15, a_2 = Order of 30, a_3 = Order of 45

$\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$
where θ_1 = Demand of 20, θ_2 = Demand of 25 and θ_3 = Demand of 30, θ_4 = Demand of 45

	a_1	a_2	a_3
θ_1	-425	-350	-50
θ_2	-400	-700	-400
θ_3	-375	-1050	-750
θ_4	-300	-975	-1800

(b) All actions are admissable.

(c) p(π,a_1) = 0.2(-425) + 0.4(-400) + 0.2(-375) + 0.2(-300)
= -380
p(π,a_2) = 0.2(-350) + 0.4(-700) + 0.2(-1050) + 0.2(-975)
= -755
p(π,a_3) = 0.2(-50) + 0.4(-400) + 0.2(-750) + 0.2(1800)
= -680
The Bayes action is a_2.

(d) $\sup_\theta(R(\theta, a_1)$ = -300
$\sup_\theta(R(\theta, a_2)$ = -350
$\sup_\theta(R(\theta, a_3)$ = -50
The minimax action is a_2 and its corresponding value is -350.

3. (a) A = {a_1: without marketing campaign, a_2: with marketing campaign }
Θ = [0,1]

$$\text{L}(\theta, a_1) = \begin{cases} -(2+3\theta), & 0 \leq \theta < 0.7 \\ -5, & 0.7 \leq \theta \leq 1 \end{cases}$$

$$\text{L}(\theta, a_2) = \begin{cases} -(4+1.5\theta), & 0 \leq \theta < 0.7 \\ -4, & 0.7 \leq \theta \leq 1 \end{cases}$$

(b) $\rho(\pi, a_1) = E^{\pi}[L(\theta, a_1)]$
$= \int_0^1 L(\theta, a_1)\pi(\theta)d\theta$
$= \int_0^{0.7} -(2+3\theta)d\theta + \int_{0.7}^1 (-5)d\theta$

= -2.135 - 1.5
= -3.635

$\rho(\pi, a_2) = E^{\pi}[L(\theta, a_2]$
$= \int_0^1 L(\theta, a_2)\pi(\theta)d\theta$
$= \int_0^{0.7} -(4 + 1.5\theta)d\theta + \int_{0.7}^1 (-4)d\theta$
= -3.1675 - 1.2
= -4.3675

Hence, a_2 is the Bayesian action.

(c) $\sup_\theta(R(\theta, a_1) = \sup_\theta(L(\theta, a_1)$ = -2 (when $\theta = 0$)
$\sup_\theta(R(\theta, a_2) = \sup_\theta(L(\theta, a_2)$ = -4 (when $\theta = 0$)
Hence, a_2 is the minimax action.

4 Minimax Analysis II (Karthika Nair and Yao-Hung Tsai)

1. For any other estimator δ', we have

$$\begin{aligned}\sup_\theta R(\theta,\delta') &\geq \int R(\theta,\delta')\pi(d\theta)\\ &\geq \int R(\theta,\delta_0^*)\pi(d\theta) = \sup_\theta R(\theta,\delta_0^*).\end{aligned}$$

   (2 pts)

   Therefore, δ_0^* is minimax. If δ_0^* si unique Bayes under π, then the second inequality is strict and thus makes δ_0^* to be a unique minimax. (1 pt)

   Let π' be another prior over Θ. Then we have

$$\begin{aligned}R(\pi',\delta_{\pi'}) &= \int R(\theta,\delta_{\pi'})\pi'(d\theta)\\ &\leq \int R(\theta,\delta_0^*)\pi'(d\theta)\\ &\leq \sup_\theta R(\theta,\delta_0^*) = R(\pi,\delta_0^*)\end{aligned}$$

   (2 pt)

   Therefore, π is least favorable. (1 pt)

2. Let δ_0 be an equalizer rule with $R(\theta,\delta_0) = C$. If δ_0 was not minimax, $\exists\delta'$ s.t. $\sup_\theta R(\theta,\delta') = v$ would be strictly less than C. (2 pts)

   For any $\epsilon > 0$ s.t. $0 < \epsilon < c - v$ and any prior π, we have

$$r(\pi,\delta_0) = c > v + \epsilon \geq r(\pi,\delta') + \epsilon.$$

   (2 pts)

   Thus, δ_0 cannot be ϵ-Bayes w.r.t. to any prior distribution, and so δ_0 can not be an extended Bayes rule. We prove this by contradiction. (2 pts)

(a)

$$
\begin{aligned}
R(\theta,\delta) =& E_X\Big(\frac{(\theta-\delta(X))^2}{1-\theta}\Big)\\
=& \sum_{x=0}^{\infty}(\theta^2-2\theta\delta(x)+\delta(x)^2)\theta^x\\
=& \delta(0)^2+\theta(\delta(1)^2-2\delta(0))+\sum_{x=2}^{\infty}(\delta(x)^2-2\delta(x-1)+1)\theta^x
\end{aligned}
$$

(4 pts)

(b)

$$
\begin{aligned}
&\delta(1)^2 = 2\delta(0) \text{ and}\\
&\delta(x)^2 = 2\delta(x-1)-1 \text{ for } x=2,3,\cdots
\end{aligned}
$$

(2 pts)

Since $2\delta(x-1)=\delta(x)^2+1=(\delta(x)-1)^2+2\delta(x)$, so that $\delta(x)$ are non-decreasing from $x=1$ on. (1 pt)

Therefore, $\delta(x)$ converges, and suppose its value is c we have $c^2=2c-1$. This implies $c=1$. (1 pt)

To conclude, we have

$$
\begin{aligned}
&\delta(0) = \frac{1}{2} \text{ and}\\
&\delta(x) = 1 \text{ for } x=1,2,3,\cdots
\end{aligned}
$$

(2 pts)

(c) From part (i),

$$
\begin{aligned}
r(\pi,\delta) =& ER(\theta,\delta)\\
=& \delta(0)^2+\mu_1(\delta(1)^2-2\delta(0))+\sum_{x=2}^{\infty}\mu_x(\delta(x)^2-2\delta(x-1)+1).
\end{aligned}
$$

(2 pts)

To find the Bayes rule, we take the derivative w.r.t. to each $\delta(x)$ for the above equation, setting to zero. (2 pts)

We then obtain

$$
\delta_\pi(x)=\frac{m_{x+1}}{m_x}
$$

for $x=0,1,2,\cdots$. (1 pt)