Homework 1
Bayesian Analysis, Minimax Analysis

CMU 10-716: Advanced Machine Learning (Spring 2019)

OUT: Jan. 24, 2019
DUE: **Feb. 5, 2019, 11:59 PM**.

Instructions:

- **Collaboration policy:** Collaboration on solving the homework is allowed, after you have thought about the problems on your own. It is also OK to get clarification (but not solutions) from books or online resources, again after you have thought about the problems on your own. There are two requirements: first, cite your collaborators fully and completely (e.g., "Bob explained to me what is asked in Question 4.3"). Second, write your solution *independently*: close the book and all of your notes, and send collaborators out of the room, so that the solution comes from you only.
- **Submitting your work:** Assignments should be submitted as PDFs using Gradescope unless explicitly stated otherwise. Each derivation/proof should be completed on a separate page. Submissions can be handwritten, but should be labeled and clearly legible. Else, submission can be written in LaTeX. Upon submission, label each question using the tempate provided by Gradescope.

1 Bayesian Analysis I (Shawn Lyu)

1. [**13 pts**] Show that for arbitrary action a and parameter θ, the Bayes rule for the weighted squared loss function $L(\theta, a) = w(\theta)(\theta - a)^2$ is

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

2. A large shipment of parts is received, out of which five are tested for defects. The number of defective parts, X, is assumed to have a $B(5, \theta)$ distribution (a binomial distribution with parameters 5 and θ). From past shipments, it is known that θ has a $Beta(3, 9)$ prior distribution.

   You may use the results from the assigned reading material as long as they are properly cited.

   (a) [**3 pts**] Derive the Bayes estimate of θ under loss

   $$L(\theta, a) = (\theta - a)^2.$$

   (b) [**3 pts**] Derive the Bayes estimate of θ under loss

   $$L(\theta, a) = |\theta - a|.$$

   For this problem, you may use the following approximation: the median of $Beta(\alpha, \beta)$ is approximately

   $$\frac{\alpha - 1/3}{\alpha + \beta - 2/3}$$

   when $\alpha, \beta > 1$.

   (c) [**6 pts**] Derive the Bayes estimate of θ under loss

   $$L(\theta, a) = (\theta - a)^2/\theta(1 - \theta).$$

2 Bayesian Analysis II (Shawn Lyu)

1. [**15 pts**] Let $\mathbf{X} = (X_1, X_2)^T \in \mathbb{R}^2$ where $\mathbf{X} \sim \mathcal{N}_2(\boldsymbol{\theta}, I_2)$ (i.e. $\mathbf{X}$ is a random vector drawn from a 2-dimensional Gaussian distribution), $L(\boldsymbol{\theta}, \mathbf{a}) = (\boldsymbol{\theta}^T\mathbf{a} - 1)^2$. Let

$$\mathcal{A} = \left\{(a_1, a_2)^T : a_1 + a_2 = 1\right\}$$

   and $\boldsymbol{\theta} = (\theta_1, \theta_2)^T \sim \mathcal{N}_2(\boldsymbol{\mu}, I_2)$ for some known parameter $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$. Find the Bayes estimator of $\boldsymbol{\theta}$.

2. [**10 pts**] Show that the Bayes estimator for estimating a vector $\boldsymbol{\theta} \in \mathbb{R}^p$ from some action $\mathbf{a} \in \mathbb{R}^p$ under a quadratic loss

$$L(\boldsymbol{\theta}, \mathbf{a}) = (\boldsymbol{\theta} - \mathbf{a})^T\mathbf{Q}(\boldsymbol{\theta} - \mathbf{a})$$

   where $\mathbf{Q}$ is a $p \times p$ positive definite matrix, is

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

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

1. [**5 pts**] Show that if a decision rule δ is admissible and has constant risk, then it is minimax. Hint: You can prove this by enumerate an arbitrary decision rule δ' and then compare their risks by the properties of admissibility and constant risk.

2. The Weasley Twins need to order their most selling product Extendable Ears for the upcoming year. Orders must be placed in quantities of 15 pairs. The cost per pair is \$40 if 15 are ordered, \$35 if 30 are ordered, and \$30 if 45 are ordered. The ears will be sold at \$70 per pair. If the Weasleys run out of ears mid year, they will suffer a loss due to unsatisfied customers. This loss is estimated at \$5 per unsatisfied customer. The Weasleys estimate the demand will be 20, 25, 30, or 45 pair of ears, with probabilities 0.2, 0.4, 0.2, and 0.2, respectively.

   (a) [**8 pts**] Describe A,Θ and the loss matrix.

   (b) [**2 pts**] Which actions are admissable?

   (c) [**2 pts**] What is the Bayes action?

   (d) [**2 pts**] What is the minimax action and the corresponding value?

3. The grim atmosphere at wizarding world sales have led to a plummet in the sales at the Weasleys' store. The twins have to now decide whether they should implement a 1 million dollar marketing strategy. If 0.7 of the inventory is sold, they believe they will make a profit of 5 million irrespective of whether the strategy was implemented or not. Let θ denote the ratio of inventory sold. If $\theta < 0.7$, they estimate their revenues to be $2 + 3\theta$ without the campaign and $4 + 1.5\theta$ with the campaign. Consider θ to have a Uniform(0,1) Distribution.

   (a) [**3 pts**] Describe A, Θ and L(θ, a).

   (b) [**3 pts**] What is the Bayes action?

   (c) [**3 pts**] What is the minimax action?

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

1. (i) [**6 pts**] Prove that if an equalizer rule δ_0^* is Bayes w.r.t. π, then δ_0^* is minimax and π is least favorable. Also explain the case if the δ_0^* is unique Bayes w.r.t. π. Hint: You can prove this by enumerate an arbitrary decision rule δ' and an arbitrary prior π' over Θ.

   (ii) [**6 pts**] Sometimes the Bayes rule may not exist because the infimum is not attained for any decision rule δ. In these cases, if for any $\epsilon > 0$, we can find a decision rule δ_ϵ for which

   $$R(\pi, \delta_\epsilon) < \inf_{\delta'} R(\pi, \delta') + \epsilon,$$

   δ_ϵ is said to be ϵ-Bayes w.r.t. π. Furthermore, a decision rule δ is said to be extended Bayes if, $\forall \epsilon > 0$, we have δ is ϵ-Bayes w.r.t. some prior (no need to be the same for different ϵ). Please show that if an equalizer rule is extended Bayes, it is a minimax rule. Hint: Prove by contradiction. An equalizer rule which is not minimax cannot be extended Bayes.

2. We have $\Theta = [0, 1)$, $\mathscr{A} = [0, 1]$, the loss function

$$L(\theta, a) = \frac{(\theta - a)^2}{(1 - \theta)},$$

and the conditional density

$$f(x|\theta) = (1 - \theta)\theta^x, \ x \in \{0, 1, 2, 3, ...\}.$$

Answer the following questions.

(i) [**2 pts**] Write down the risk function $R(\theta, \delta)$ for $\delta \in \mathscr{D}$.

(ii) [**6 pts**] Show that there exist an unique equalizer rule and write down the rule. Hint: Recall the definition of the equalizer rule anda void it to be infinity.

(iii) [**5 pts**] Show that $\delta \in \mathscr{D}$ is Bayes w.r.t. π *iff* $\delta(k) = \frac{m_{k+1}}{m_k}$ (k being positive integers) with m_k are the moments of the distribution of π.