Hw1

2022年2月25日 下午 1:31

Problem 1: 1. Suppose: $h(x)$ means the x-time draws heart.

$P_1 = P(h(2)\mid h(1)) = P(h(2)) = \frac{1}{4}$.

Since the first and second draw are independent

2.

2 \ 1	heart	others
heart	$\frac{1}{4}\times\frac{1}{4}$	$\frac{3}{4}\times\frac{1}{4}$
others	$\frac{1}{4}\times\frac{3}{4}$	$\frac{3}{4}\times\frac{3}{4}$

$\therefore P_2 = \dfrac{\frac{1}{4}\times\frac{1}{4}}{\frac{1}{4}\times\frac{3}{4}+\frac{1}{4}\times\frac{3}{4}+\frac{1}{4}\times\frac{1}{4}} = \frac{1}{7}$.

Homework 1

Machine Learning

Due on March 8 at 12:00PM (noon) on Canvas

Problem 1 (4 points)

不放回抽样.

Two cards are drawn from a deck of 52 cards without replacement.

1. What is the probability that the second card is a heart, given that the first card is a heart?
2. What is the probability that both cards are hearts, given that at least one card is a heart?

Problem 2 (6 points)

One card is selected from a deck of 52 cards and placed in a second deck containing 52 cards. A card is then selected from the second deck.

1. What is the probability that a card drawn from the second deck is an ace?
2. If the first card is placed into a deck of 54 cards containing two jokers, then what is the probability that a card drawn from the second deck is an ace?
3. Given that an ace was drawn from the second deck in (ii), what is the conditional probability that an ace was transferred from the first deck?

Card picked from deck 1 is ace; picked card is not ace.

Problem 2. 1. Solve: $P = \frac{1}{13}\times\frac{5}{53} + \frac{12}{13}\times\frac{1}{13} = \frac{53}{689}$

2. Solve: $P(Ace) = \frac{1}{13}\times\frac{5}{55} + \frac{12}{13}\times\frac{4}{55} = \frac{53}{715}$

3. Solve: $P = \dfrac{P(\text{Ace from first deck})}{P(Ace)} = \dfrac{\frac{1}{13}\times\frac{5}{55}}{\frac{53}{715}} = \frac{5}{53}$.

12/21 15/51/24/42/33, 3, 6, 9 → 36/63/45/54

6×6=36

Problem 3 (3 points)

You and your friend are playing the following game: two dice are rolled; if the total showing is divisible by 3, you pay your friend \$6. If you want to make the game fair, how much should she pay you when the total is not divisible by 3? A fair game is one in which your expected winnings are \$0.

Problem 3: $P(S\equiv 0 \pmod 3) = \frac{1}{6}\times\frac{1}{3}+\frac{1}{6}\times\frac{1}{6} = \frac{1}{18}$.

$\therefore E(\text{winning}) = \frac{1}{18}\times 6\$ - \frac{17}{18}x = 0 \Rightarrow x = \frac{6}{17}\$$

Problem 4 (6 points)

Suppose we have n samples $x_1, x_2, \ldots, x_n$ independently drawn from a normal distribution with *known* variance σ^2 and *unknown* mean μ.

Problem 4: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

1. $f(x_1, x_2, \ldots, x_n \mid \mu) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$

$\ln f = -n\ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$

$(\ln f(\mu))' = +\frac{1}{2\sigma^2}\cdot 2\sum_{i=1}^{n}(x_i-\mu)$

$= \dfrac{\sum_{i=1}^{n}x_i - n\mu}{\sigma^2}$ let the derivation equals to 0.

$\Rightarrow n\mu = \sum_{i=1}^{n} x_i \Rightarrow \mu = \dfrac{\sum_{i=1}^{n} x_i}{n}$

2. $p(\mu) = \frac{1}{\beta\sqrt{2\pi}} e^{-\frac{(\mu-\gamma)^2}{2\beta^2}}$

$\therefore p(\mu \mid x_1, x_2, \ldots, x_n) \propto p(\mu)p(x_1, x_2, \ldots, x_n \mid \mu)$

$\ln(p(\mu)p(x\mid\mu)) = -\ln(\beta\sqrt{2\pi}) - \frac{(\mu-\gamma)^2}{2\beta^2} - n\ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$

derivate μ on both side, let it equals to zero.

$\ln'(p(\mu)p(x\mid\mu)) = -\frac{1}{2\beta^2}2(\mu-\gamma) + \frac{1}{2\sigma^2}2\sum_{i=1}^{n}(x_i-\mu)$.

$\frac{\mu-\gamma}{\beta^2} = \frac{\sum_{i=1}^{n}x_i - n\mu}{\sigma^2} \Rightarrow \mu' = \frac{\sigma^2}{n\beta^2+\sigma^2}\gamma + \frac{\beta^2}{n\beta^2+\sigma^2}\sum_{i=1}^{n}x_i$

1. Please derive the MLE estimator for the mean μ.
2. Assume the prior distribution for the mean is a normal distribution with mean γ and variance β^2. Please derive the MAP estimator for the mean μ.
3. Please comment on the MLE and MAP estimators as the number of samples n goes to infinity.

3. when $n\to\infty$, $\frac{\beta^2}{n\beta^2+\sigma^2} \to \frac{1}{n}$. $\mu' \to \sum_{i=1}^{n} x_i$ (MAP) which equals to μ (MLE)

Problem 5 (2 points)

Let $\mathbf{A} = \begin{bmatrix} 1 & -1 & 1 \\ 2 & 0 & 1 \\ 3 & 0 & 2 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 2 & -2 \\ 1 & 3 \\ 5 & -4 \end{bmatrix}$, please verify that $A(AB) = A^2B$.

Problem 5.

left: $AB = \begin{bmatrix} 6 & -9 \\ 9 & -8 \\ 16 & -14 \end{bmatrix}$ $A(AB) = \begin{bmatrix} 13 & -15 \\ 28 & -32 \\ 50 & -55 \end{bmatrix}$

right: $A^2 = \begin{bmatrix} 2 & -1 & 2 \\ 5 & -2 & 4 \\ 9 & -3 & 7 \end{bmatrix}$ $A^2B = \begin{bmatrix} 13 & -15 \\ 28 & -32 \\ 50 & -55 \end{bmatrix}$

Problem 6 (6 points)

Let $f(\mathbf{w}) = (\mathbf{y} - \mathbf{X}\mathbf{w})^\top(\mathbf{y} - \mathbf{X}\mathbf{w}) + \frac{1}{2}\mathbf{w}^\top\mathbf{w}$ where $\mathbf{y}, \mathbf{a} \in \mathbb{R}^{m\times 1}$, $\mathbf{w} \in \mathbb{R}^{n\times 1}$, and $\mathbf{X} \in \mathbb{R}^{m\times n}$.

$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \end{bmatrix}_m \begin{bmatrix} \; \end{bmatrix}_n$

1. What is the **gradient** of $f(\mathbf{w})$, $\nabla f(\mathbf{w})$?
2. What is the **Hessian** of $f(\mathbf{w})$, $\nabla^2 f(\mathbf{w})$?
3. Can we conclude that $f(\mathbf{w})$ is a convex function? Please explain.

Problem 6. $y - Xw = \begin{bmatrix} y_1 - \sum_{i=1}^{n} x_{1i}\cdot w_i \\ \vdots \\ y_m - \sum_{i=1}^{n} x_{mi} w_i \end{bmatrix}_{m\times 1}$

$\therefore f(w) = \sum_{j=1}^{m}\left(y_j - \sum_{i=1}^{n} x_{ji}\cdot w_i\right)^2 + \frac{1}{2}\sum_{i=1}^{n} w_i^2$ is a scalar

1. $\nabla f(w)_{n\times 1} = \begin{bmatrix} \frac{\partial f(w)}{\partial w_1} \\ \vdots \\ \frac{\partial f(w)}{\partial w_n} \end{bmatrix} = \begin{bmatrix} -2\sum_{j=1}^{m} x_{j1}(y_j - x_{j1}w_1) + w_1 \\ \vdots \\ -2\sum_{j=1}^{m} x_{jn}(y_j - x_{jn}w_n) + w_n \end{bmatrix} = -2X^{T}_{n\times m}(y - Xw)_{m\times 1} + w$.

2. $\nabla^2 f(w)_{n\times n} = \begin{bmatrix} \frac{\partial f(w)}{\partial w_1 \partial w_1} & \cdots & \frac{\partial f(w)}{\partial w_1 \partial w_n} \\ \vdots & & \end{bmatrix} = \begin{bmatrix} 2\sum_{j=1}^{m} x_{j1}^2 + 1 & \cdots & 0 \\ \vdots & & \vdots \end{bmatrix}$

Problem 7 (3 points)

Prove that the intersection of multiple convex sets is a convex set. In other words, if $C_i, i = 1, 2, \ldots, n$ are convex sets, please show that the intersection, $C = \cap_{i=1}^{n} C_i$, is a convex set.

Problem 8 (6 points)

1. Let $\mathrm{dom}(\bar f)$ denote the domain of the function $\bar f$. Given a set of convex functions $\{f_i : i = 1, \ldots, n\}$ and a set of nonnegative weights $\{w_i : i = 1, \ldots, n\}$, prove that $f = \sum_{i=1}^{n} w_i f_i$ is a convex function on its domain $\mathrm{dom}(f) = \cap_{i=1,2,\ldots,n}\mathrm{dom}(f_i)$.
2. Let Y denote a discrete random variable that can take possibles values from the set $\{y_1, y_2, \ldots, y_n\}$.

2. $\nabla^2 f(w)_{n\times n} = \begin{bmatrix} \frac{\partial f(w)}{\partial w_1 \partial w_1} & \cdots & \frac{\partial f(w)}{\partial w_1 \partial w_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(w)}{\partial w_n \partial w_1} & \cdots & \frac{\partial f(w)}{\partial w_n \partial w_n} \end{bmatrix} = \begin{bmatrix} 2\sum_{j=1}^{m} x_{j1}^2 + 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 2\sum_{j=1}^{m} x_{jn}^2 + 1 \end{bmatrix}$

$= 2X^TX + I$

3. Since $\nabla^2 f(w)$ is positive definite, that is for any vector $y \in R^n$ and $y \neq 0$, there is $y^T \nabla^2 f(w) y = \sum_{i=1}^{n} 2y_i^2 x_i^2 + y_i^2 > 0$. $\therefore \nabla^2 f(w) > 0$.
therefore $f(w)$ is a convex function.

1. Let dom(f) denote the domain of the function f. Given a set of convex functions $\{f_i : i =$ $1, \ldots, n\}$ and a set of nonnegative weights $\{w_i : i = 1, \ldots, n\}$, prove that $f = \sum_{i=1}^{n} w_i f_i$ is a convex function on its domain $\mathrm{dom}(f) = \cap_{i=1,2\ldots,n}\mathrm{dom}(f_i)$.

2. Let Y denote a discrete random variable that can take possibles values from the set $\{y_1, y_2, \ldots, y_n\}$. Suppose $f(x, y)$ is convex in $x \in \mathbb{R}$ for any fixed y, please show that $g(x) = \mathrm{E}_Y[f(x, Y)]$ is convex in x.

Problem 7

Here we take two convex set C_1, C_2 to prove.
for any n greater than 2, it's just the same.

for $x, y \in C_1 \cap C_2$. and any $\alpha \in [0,1]$

① $x, y \in C_1$: $\alpha x + (1-\alpha) y \in C_1$

② $x, y \in C_2$: $\alpha x + (1-\alpha) y \in C_2$

therefore: $\alpha x + (1-\alpha) y \in C_1 \cap C_2$

And according to the definition of convex set,
$C = C_1 \cap C_2$ is a convex set.

Problem 8.

1. We also take two convex function f_1, f_2 as example.
for $x, y \in \mathrm{dom}(f_1) \cap \mathrm{dom}(f_2)$, and $\alpha \in [0,1]$, $w_1, w_2 \geq 0$, we have:
① f_1 is a convex function: $f_1(\alpha x + (1-\alpha)y) \leq \alpha f_1(x) + (1-\alpha) f_1(y)$.
② f_2 is a convex function: $f_2(\alpha x + (1-\alpha)y) \leq \alpha f_2(x) + (1-\alpha) f_2(y)$.
$\because f = w_1 f_1 + w_2 f_2$. $\therefore f(\alpha x + (1-\alpha)y) = w_1 f_1(\alpha x + (1-\alpha)y) + w_2 f_2(\alpha x + (1-\alpha)y)$
$\leq w_1[\alpha f_1(x) + (1-\alpha) f_1(y)] + w_2[\alpha f_2(x) + (1-\alpha) f_2(y)]$
$= \alpha[w_1 f_1(x) + w_2 f_2(x)] + (1-\alpha)[w_1 f_1(y) + w_2 f_2(y)]$
$= \alpha f(x) + (1-\alpha) f(y)$. Q.E.D.

2. proof: for any $x_1, x_2 \in \mathrm{dom}(g)$, and any $\alpha \in [0,1]$.
① $f(x,y)$ is convex: $f(\alpha x_1 + (1-\alpha)x_2, y) \leq \alpha f(x_1, y) + (1-\alpha) f(x_2, y)$ ①
$g(\alpha x_1 + (1-\alpha)x_2) = E_Y[f(\alpha x_1 + (1-\alpha)x_2, Y)]$
$= \frac{1}{n}\sum_{i=1}^{n}[f(\alpha x_1 + (1-\alpha)x_2, y_i)] \leq \frac{1}{n}\sum_{i=1}^{n}[\alpha f(x_1, y_i) + (1-\alpha) f(x_2, y_i)]$
$= \frac{1}{n}\sum_{i=1}^{n}\alpha f(x_1, y_i) + \frac{1}{n}\sum_{i=1}^{n}(1-\alpha) f(x_2, y_i) = \alpha E_Y[f(x_1, Y)] + (1-\alpha) E_Y[f(x_2, Y)] = \alpha g(x_1) + (1-\alpha) g(x_2)$. Q.E.D.