440/540 Artificial Intelligence HW8

INSTRUCTIONS

- **Due: Wednesday, 19 April 2023 at 23:59 PM PDT** Remember that you have 5 slip days to use at your discretion BUT you can use no more than 1 per homework.
- **Format:** Submit the answer sheet containing your answers in PDF. You can provide either typed or handwritten answers to this homework. Handwritten answers must be legible and scanned into a PDF.
- **Note:** **Please DO NOT FORGET to include your name and WSU ID in your submission.**
- **How to submit:** Submit a PDF containing your answers on Canvas
- **Policy:** See the course website for homework policies and Academic Integrity.

Last Name	Ghosh Dastidar
First Name	Priyanka
WSU ID	11804812

Q1. [40 pts] Maximum Likelihood Estimation (I)

Given n = 100 (random) observations $x_1,x_2,\ldots,x_n$ which are independently drawn from an univariate Gaussian distribution N($2\mu, 7\sigma^2$) with unknown mean μ and variance $\sigma^2 > 0$.

(a) [20 pts] Derive the maximum likelihood estimation μ_{MLE} of μ as a function of n and ($x_1,x_2,\ldots,x_n$).

Answer:

Since the observations are independently drawn from a Gaussian distribution with mean 2μ and variance 7 σ^2, the likelihood function can be written as:

L(μ|x1, x2, . . . , xn) = f(x1, x2, . . . , xn|μ)
= (1 / √(2π * 7σ^2))^n * exp(-(1/2*7σ^2) * ∑(xi - 2μ)^2)

To maximize the likelihood function, the derivative of the log likelihood function with respect to μ is set equal to 0:

d/dμ log(L(μ|x1, x2, . . . , xn)) = d/dμ [nlog(1/√(2π7σ^2)) - ∑(xi-2μ)^2/(27σ^2)]
= d/dμ(nlog(1 / √(2π * 7σ^2)) - (1/2*7σ^2) * ∑(xi - 2μ)^2)
= -1/(7σ^2) * ∑(xi-2μ)

Setting the derivative to 0 and solving for μ, we get:
1/7σ^2 * ∑(xi - 2μ) = 0
∑(xi - 2μ) = 0
∑xi - 2nμ = 0

μMLE = (1/n) * ∑xi / 2

Therefore, the MLE of μ is equal to the average of the observations divided by 2.

(b) [20 pts] Derive the maximum likelihood estimation σ_{MLE}^2 of σ^2 as a function of n and ($x_1,x_2\ldots,x_n$).

Answer:

To derive the maximum likelihood estimation (MLE) of σ^2, should find the value of σ^2 that maximizes the likelihood function L(σ^2|x1, x2, . . . , xn).

Using the same Gaussian distribution as in part (a), the likelihood function can be written as:
L(σ^2|x1, x2, . . . , xn) = f(x1, x2, . . . , xn|σ^2)
= (1/√(2π7σ^2))^n * exp(-∑(xi-2μ)^2/(2*7σ^2))

To maximize the likelihood function, we can take the derivative of the log-likelihood function with respect to σ^2 and set it equal to 0:
d/dσ^2 log(L(σ^2|x1, x2, . . . , xn))
= d/dσ^2 [nlog(1/√(2π7σ^2)) - ∑(xi-2μ)^2/(27σ^2)]

$= -n/(2\sigma^2) + 1/(14\sigma^4) * \sum(x_i-2\mu)^2$

Setting the derivative to 0 and solving for σ^2, we get:
$\sigma^2_{MLE} = (1/n) * \sum(x_i-2\mu)^2/7$

Therefore, the MLE of σ^2 is equal to the average of the squared differences between each observation and 2μ, divided by 7.

Q2. [30 pts] Maximum Likelihood Estimation (II)

Let $x_1,x_2,...,x_n$ denote the n independent observations which are assumed to be drawn from the same distribution $p(x \mid \theta)$ with defining parameter θ.

(a) [10 pts] Suppose $0 < \theta < 1$ and $p(x = 1 \mid \theta) = \theta$ while $p(x = 0 \mid \theta) = 1 - \theta$. Then, suppose m out of n observations ($m < n$) have value 1 while the rest has value 0.

Compute the maximum likelihood estimation θ_{MLE} of θ in terms of m and n.

Answer:

To find the maximum likelihood estimation (MLE) of θ, the value of θ that maximizes the likelihood function $L(\theta \mid x1, x2, . . . , xn)$ is to be found.

The likelihood function can be written as:
$L(\theta \mid x1, x2, . . . , xn) = p(x1, x2, . . . , xn \mid \theta) = \theta^m * (1-\theta)^{(n-m)}$

To maximize the likelihood function, the derivative of the log-likelihood function is taken with respect to θ and set it equal to 0:
$d/d\theta \log(L(\theta \mid x1, x2, . . . , xn)) = d/d\theta [m*\log(\theta) + (n-m)*\log(1-\theta)]$
$= m/\theta - (n-m)/(1-\theta)$

Setting the derivative to 0 and solving for θ:
$\theta_{MLE} = m/n$

Therefore, the MLE of θ is equal to the ratio of the number of observations that take the value 1 to the total number of observations.

(b) [10 pts] Suppose $\theta > 0$ and assume the observations $x_1,x_2,...,x_n$ were independently drawn from Uniform$(0,1/\theta)$. Assuming all observations are positive, show that $p(x_i \mid \theta) = \mathrm{I}(\theta \leq 1/x_i) \cdot \theta$.

Answer :

Here, $I(\theta \leq 1/x_i) = 1$ if and only if $\theta \leq 1/x_i$ is true and if $x \sim$ Uniform(a,b) then $p(x \mid a,b) = 1/(b - a)$.
The probability density function of the uniform distribution is given by:
$p(x \mid a, b) = 1/(b - a)$, for $a \leq x \leq b$

Substituting $a = 0$ and $b = 1/\theta$:

$p(x \mid \theta) = 1/(1/\theta - 0) = \theta$, for $0 \le x \le 1/\theta$

Since all observations are positive, the above probability density function holds only if $\theta \le 1/x$. Therefore:
$p(xi \mid \theta) = I(\theta \le 1/xi) \cdot \theta$

where $I(\theta \le 1/xi)$ is an indicator function which takes the value 1 if the condition $\theta \le 1/xi$ is true and 0 otherwise. This indicates that the probability density function is only valid when θ is less than or equal to the inverse of the observation xi.

Therefore, it is shown that the probability density function of each observation xi given θ is equal to θ when θ is less than or equal to the inverse of the observation xi, and 0 otherwise.

(c) [10 pts] Following the same setting in part (b), compute θ_{MLE} in terms of $(x_1,x_2,\ldots,x_n)$

Answer:

From part (b), we know that the probability density function of each observation xi given θ is:

$p(xi \mid \theta) = I(\theta \le 1/xi) \cdot \theta$

The likelihood function can be written as:

$L(\theta \mid x1, x2, \ldots, xn) = p(x1, x2, \ldots, xn \mid \theta) = \Pi i{=}1n\ I(\theta \le 1/xi) \cdot \theta$

To find the MLE of θ, we need to maximize the likelihood function with respect to θ. However, note that the indicator function $I(\theta \le 1/xi)$ makes this function non-differentiable at certain points.

We can simplify the likelihood function by noting that the maximum value of θ occurs when θ is the smallest possible value satisfying the constraint $\theta \le 1/xi$ for all i. This means that:

$\theta MLE = \min\{1/x1, 1/x2, \ldots, 1/xn\}$

Therefore, the MLE of θ is equal to the reciprocal of the largest observation.

Q3. [30 pts] Maximum a Posteriori

Let $x_1,x_2,\ldots,x_n$ denote the $n > 2$ independent observations which are assumed to be drawn from the same distribution $p(x \mid \theta)$ where $\theta \sim \mathrm{Beta}(a,b)$ with $a,b > 0$.

The probability density function (PDF) of the beta distribution is $p(\theta \mid a,b) = \theta^{a-1}(1-\theta)^{b-1}/\mathrm{B}(a,b)$ where $\mathrm{B}(a,b) = \mathrm{G}(a)\mathrm{G}(b)/\mathrm{G}(a+b)$ with G denote the Gamma function as mentioned in slides 46-47 of lecture 22. This density is non-zero only at $\theta \in (0,1)$.

(a) [15 pts] Assume $p(x \mid \theta)$ is the Bernoulli distribution as in part (a) of Q2. That is, $p(x = 1 \mid \theta) = \theta$ while $p(x = 0 \mid \theta) = 1 - \theta$ with $0 < \theta < 1$.

Given that m out of n observations $(1 < m < n)$ has value 1, derive θ_{MAP} in terms of m, n, a and b.

Answer:

The prior distribution of θ is Beta(a, b), so the posterior distribution of θ given the observations can be calculated as follows:
p(θ | x1, x2, ..., xn) ∝ p(x1, x2, ..., xn | θ) p(θ)
∝ θ^m (1-θ)^(n-m) θ^(a-1) (1-θ)^(b-1)
∝ θ^(m+a-1) (1-θ)^(n-m+b-1)

The posterior distribution is again a Beta distribution with parameters m+a and n-m+b, which means that the posterior mean is (m+a)/(n+a+b) and the posterior mode is (m+a-1)/(n+a+b-2).

Therefore, the maximum a posteriori (MAP) estimate of θ is:
θMAP = (m+a-1)/(n+a+b-2)

Substituting the given values of the Bernoulli distribution:
θMAP = (m+a-1)/(n+a+b-2)

(b) [15 pts] Now, assume instead that each observation can take on 3 values in {0,1,2} according to the following distribution: $p(x = 0 \mid \theta) = \theta$, $p(x = 1 \mid \theta) = \theta \cdot (1 - \theta)$, and $p(x = 2 \mid \theta) = (1 - \theta)^2$ with $0 < \theta < 1$.

Given that there are n_0, n_1 and n_2 observations with values 0, 1 and 2, respectively. Derive θ_{MAP} in terms of n_0, n_1, n_2, a and b if $n_0, n_1, n_2 \geq 1$.

Answer:

The likelihood function in this case is:
p(x | θ) = θ^(n0) * θ^(n1) * (1-θ)^(n1) * (1-θ)^(2n2)
= θ^(n0+n1) * (1-θ)^(n1+2n2)

Using the same prior distribution as before, the posterior distribution is:
p(θ | x, a, b) ∝ p(x | θ) p(θ | a, b)
∝ θ^(n0+n1+a-1) * (1-θ)^(n1+2n2+b-1)

Taking the derivative of the log of the posterior with respect to θ:
d/dθ [log p(θ | x, a, b)]= d/dθ [(n0+n1+a-1) log θ + (n1+2n2+b-1) log (1-θ)]
= (n0+n1+a-1)/θ - (n1+2n2+b-1)/(1-θ)

Setting the above equal to 0:
(n0+n1+a-1)/θ = (n1+2n2+b-1)/(1-θ)
(n0+n1+a-1)(1-θ) = θ(n1+2n2+b-1)
(n0+n1+a-1) - θ(n0+n1+a-1) = θ(n1+2n2+b-1)
θ = (n0+n1+a-1)/(n0+2n1+2n2+a+b-2)

Therefore, the MAP estimate of θ in terms of n0, n1, n2, a, and b is:
θMAP = (n0+n1+a-1)/(n0+2n1+2n2+a+b-2)