$[\text{Omar Ali Ramadan}, \text{Jason Cramer}]^{\top}$

CS189 $\qquad$ Homework 2

Problem 1. A target is made of 3 concentric circles of radii $1/\sqrt{3}$, 1 and $\sqrt{3}$ feet. Shots within the inner circle are given 4 points, shots within the next ring are given 3 points, and shots within the third ring are given 2 points. Shots outside the target are given 0 points.

Let X be the distance of the hit from the center (in feet), and let the p.d.f of X be

$$f(x) = \begin{cases} \frac{2}{\pi(1+x^2)} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

What is the expected value of the score of a single shot?

Solution. Define our score to be a function $g(x)$

$$g(x) = \begin{cases} 4 & 0 \leq x < 1/\sqrt{3} \\ 3 & 1/\sqrt{3} \leq x < 1 \\ 2 & 1 \leq x < \sqrt{3} \\ 0 & \text{otherwise} \end{cases}$$

The expected value of the score of a single shot is

$E[x] = \displaystyle\int_{-\infty}^{\infty} g(x) f(x)\, dx.$

$E[x] = \displaystyle\int_{0}^{1/\sqrt{3}} 4 * \frac{2}{\pi(1+x^2)}\, dx + \int_{1/\sqrt{3}}^{1} 3 * \frac{2}{\pi(1+x^2)}\, dx + \int_{1}^{\sqrt{3}} 2 * \frac{2}{\pi(1+x^2)}\, dx.$

$E[x] = 4 * \left. \dfrac{2 * tan^1(x)}{\pi} \right|_{0}^{1/\sqrt{3}} + 3 * \left. \dfrac{2 * tan^1(x)}{\pi} \right|_{1/\sqrt{3}}^{1} + 2 * \left. \dfrac{2 * tan^1(x)}{\pi} \right|_{1}^{\sqrt{3}}$

$E[x] = 4/3 + 1/2 + 1/3$

$E[x] = 13/6$

$\square$

Problem 2. Assume that the random variable X has the exponential distribution

$$f(x|\theta) = \theta e^{-\theta x} \qquad\qquad x > 0, \theta > 0$$

where θ is the parameter of the distribution. Use the method of maximum likelihood to estimate θ if 5 observations of X are $x_1 = 0.9$, $x_2 = 1.7$, $x_3 = 0.4$, $x_4 = 0.3$, and $x_5 = 2.4$.

Solution. Define $l(x|\theta)$ to be the log-likelihood

$$l(x|\theta) = log(\theta) - \theta x$$

To maximize the likelihood estimate of θ, maximize the sum of the log likelihood $l(x|\theta)$

$$\frac{\partial}{\partial\theta} 5log(\theta) - \theta(0.9 + 1.7 + 0.4 + 0.3 + 2.4) = 0$$

$$\frac{5}{\theta} - 5.7 = 0$$

$$\theta = \frac{5}{5.7} = 0.877$$

□

Problem 3. Let X have a Laplace distribution with density

$$f(x|\mu, b) = \frac{1}{2b} exp(-\frac{|x - \mu|}{b})$$

Suppose that n samples $x_1 \ldots x_n$ are drawn independently according to $f(x|\mu, b)$.

(a) Find the maximum likelihood estimate of μ.

Solution. Define $l(x|\theta)$ to be the log-likelihood

$$l(x|\mu, b) = log(\frac{1}{2b}) - \frac{|x - \mu|}{b}$$

To maximize the likelihood estimate of μ, maximize the sum of the log likelihood $l(x|\theta)$

$$\frac{\partial(nlog(\frac{1}{2b}) - \frac{1}{b}\sum_{i=1}^{n}|x_i - \mu|)}{\partial\mu} = 0$$

$$0 - \frac{1}{b}\sum_{i=1}^{n}\frac{\mu - x_i}{|x_i - \mu|} = 0$$

$$\frac{1}{b}\sum_{i=1}^{n}\frac{\mu - x_i}{|x_i - \mu|} = 0$$

$$\frac{1}{b}\sum_{i=1:x_i\leq\mu}^{n}\frac{\mu - x_i}{|x_i - \mu|} = \frac{1}{b}\sum_{i=1:x_i>\mu}^{n}\frac{x_i - \mu}{|x_i - \mu|}$$

From this, we see that the number of data points less than or equal to μ must be equal to the number of data points greater than μ. Therefore, log likelihood is maximized when μ is the median of the data points. That is,

$$\mu_{\mathrm{MLE}} = \mathrm{median}(x_1, \cdots, x_n)$$

□

(b) Find the maximum likelihood estimate of b.

Solution.

$$\frac{\partial(n\log(\frac{1}{2b}) - \frac{1}{b}\sum_{i=1}^{n}|x_i - \mu|)}{\partial b} = 0$$

$$-\frac{n}{b} + \frac{1}{b^2}\sum_{i=1}^{n}|x_i - \mu| = 0$$

$$\frac{1}{b^2}\sum_{i=1}^{n}|x_i - \mu| = \frac{n}{b}$$

$$\frac{1}{n}\sum_{i=1}^{n}|x_i - \mu| = b$$

□

(c) Show that b_{MLE} is an unbiased estimator (to show that the estimator is unbiased show that $E[b_{MLE} - b] = 0$)

Solution.

$$E[b_{MLE}] = E[\frac{1}{n}\sum_{i=1}^{n}|x_i - \mu|]$$

$$E[b_{MLE}] = \frac{1}{n}\sum_{i=1}^{n}E[|x_i - \mu|]$$

From the Wikipedia page on the Laplace distribution, we find that $|X - \mu| \sim$

$Exp(\frac{1}{b})$. Therefore,

$$E[|X - \mu|] = \frac{1}{1/b} = b$$

$$E[b_{MLE}] = \frac{1}{n}\sum_{i=1}^{n} E[|x_i - \mu|]$$

$$E[b_{MLE}] = \frac{1}{n}\sum_{i=1}^{n} b = \frac{n}{n}b = b$$

$$E[b_{MLE}] - b = b - b = 0$$

Therefore, b_{MLE} is an unbiased estimator of b. $\square$

Problem 4. Let $x = \begin{bmatrix} x_1 \cdots x_n \end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^n$, and let $A \in \mathbb{R}^{n \times n}$ be the square matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

(a) Give an explicit formula for $x^{\mathsf{T}}Ax$. Write your answer as a sum involving the elements of A and x.

> *Solution.* Let $\mathbf{a}_j$ be the j column vector of A.
>
> $$\mathbf{x}^{\mathsf{T}}A\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 \cdot \mathbf{x} \cdots \mathbf{a}_n \cdot \mathbf{x} \end{bmatrix} \mathbf{x}$$
>
> $$= \sum_{j=1}^{n}(\mathbf{a}_j \cdot \mathbf{x})x_j = \sum_{j=1}^{n}(\sum_{i=1}^{n} a_{ij}x_i)x_j$$
>
> $$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{i,j}x_i x_j$$
>
> $\square$

(b) Show that if A is positive definite, then the entries on the diagonal of A are positive (that is, $a_{ii} > 0$ for all $1 \leq i \leq n$).

> *Solution.* We will use a proof by contradiction to show that if A is a positive definite, $a_{ii} > 0$ for all $1 \leq i \leq n$.
> Assume that A is a positive definite. That is, $\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j \geq 0 \forall x$. Assume that $\exists i$ s.t $a_{ii} \leq 0$.
> Consider the case where $x_i \neq 0$ and $x_j = 0 \leq j \leq n, j \neq i$. $\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j = a_{ii}x_i x_i = a_{ii}x_i^2 \leq 0$.
> Contradiction. A is not positive definite.

Therefore no $a_{ii} \leq 0$. $a_{ii} > 0$ for all $1 \leq i \leq n$ □

Problem 5. Let B be a positive semidefinite matrix. Show that $B + \gamma I$ is positive definite for any $\gamma > 0$.

Solution. Suppose we have the quadratic form $\mathbf{x}^\intercal (B + \gamma I)\mathbf{x}$, where $\mathbf{x} =$ is some n-dimensional vector and $\mathbf{x} \neq \mathbf{0}$.

$$\mathbf{x}^\intercal (B + \gamma I)\mathbf{x} = \mathbf{x}^\intercal B\mathbf{x} + \gamma \mathbf{x}^\intercal I\mathbf{x} = \mathbf{x}^\intercal B\mathbf{x} + \gamma \mathbf{x}^\intercal \mathbf{x} = \mathbf{x}^\intercal B\mathbf{x} + \gamma \|\mathbf{x}\|^2$$

Since B is positive semidefinite, we know that

$$\mathbf{x}^\intercal B\mathbf{x} \geq 0$$

Additionally, since $\mathbf{x} \neq \mathbf{0}$ and $\gamma > 0$

$$\gamma \|\mathbf{x}\|^2 > 0$$

From these properties, we can assert that

$$\mathbf{x}^\intercal B\mathbf{x} + \gamma \|\mathbf{x}\|^2 = \mathbf{x}^\intercal (B + \gamma I)\mathbf{x} > 0$$

Therefore, by definition, $B + \gamma I$ is positive definite.   □

Problem 6. Suppose we have a classification problem with classes labeled $1, \cdots, c$ and an additional doubt category labeled as $c + 1$. Let the loss function be the following:

$$\lambda(\alpha_i | \omega_j) = \begin{cases} 0 & \text{if} i = j | i, j \in \{1, \cdots, c\} \\ \lambda_r & \text{if} i = c + 1 \\ \lambda_s & \text{otherwise} \end{cases}$$

where λ_r is the loss incurred for choosing doubt and λ_s is the loss incurred for making a misclassification.

(a) Show that the minimum risk is obtained if we follow this policy: (1) choose class i if $P(\omega_i | x) \geq P(\omega_j | x)$ for all j and $P(\omega_i | x) \geq 1 - \lambda_r / \lambda_s$, and (2) choose doubt otherwise.

Solution. The risk for this classifier is $R(\alpha_i | x) = \sum_{j=1}^{c+1} \lambda(\alpha_i | \omega_j) P(\omega_j | x)$. We

want to find the class i that will minimize the risk, that is

$$\arg\min_i R(\alpha_i|x) = \arg\min_i \{\lambda_r, \min \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j)P(\omega_j|x)\}$$

$$= \arg\min_i \{\lambda_r, \min \sum_{j=1:j\neq i}^{c} \lambda(\alpha_i|\omega_j)P(\omega_j|x)\}$$

$$= \arg\min_i \{\lambda_r, \min \sum_{j=1:j\neq i}^{c} \lambda_s P(\omega_j|x)\}$$

$$= \arg\min_i \{\lambda_r, \min \lambda_s(1 - P(\omega_i|x))\}$$

The first term in the arg min corresponds to picking doubt $(i = c+1)$, and the second corresponds to choosing a label. We pick a label when $\lambda_s(1-P(\omega_i|x)) \leq \lambda_r$, for whatever label i minimized the left quantity. From this, we get

$$(1 - P(\omega_i|x)) \leq \frac{\lambda_r}{\lambda_s}$$

$$P(\omega_i|x) \geq 1 - \frac{\lambda_r}{\lambda_s}$$

The label we pick in this case minimizes $1-P(\omega_i|x)$, which maximizes $P(\omega_i|x)$. That is, we pick the label i such that

$$P(\omega_i|x) \geq P(\omega_j|x), \forall j$$

If $P(\omega_i|x) < 1 - \frac{\lambda_r}{\lambda_s}$, we choose doubt. This is the policy that minimizes risk, which is the same as the proposed policy.   □

(b) What happens when $\lambda_r = 0$? What happens when $\lambda_r > \lambda_s$?

Solution. Let us consider the case where $\lambda_r = 0$. From this we know that $1 - \frac{\lambda_r}{\lambda_s} = 1$. If $P(\omega_i|x) \neq 1$, then $P(\omega_i|x) < 1$, so we will always choose doubt according to our policy, as there is no penalty for choosing doubt. If we can have a $P(\omega_i|x) = 1$, then we will be indifferent to doubt or choosing a label (though strictly, our policy will choose to assign a label).

Now we consider the case where $\lambda_r > \lambda_s$. From this we know that $1 - \frac{\lambda_r}{\lambda_s} < 0$. Since probability is non-negative, we know that $P(\omega_i|x) > 1 - \frac{\lambda_r}{\lambda_s}$. This means that we will never pick doubt and always pick a label, as the loss for choosing doubt is greater than mis-classification.   □

Problem 8. Recall that the probability mass function of a Poisson random variable is

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!} x \in \{0, 1, \cdots, \infty\}$$

You are given two equally likely classes of Poisson data with parameters $\lambda_1 = 10$ and $\lambda_2 = 15$. This means that $x|\omega_1 \sim \text{Poisson}(\lambda_1)$ and $x|\omega_2 \sim \text{Poisson}(\lambda_2)$.

(a) Find the optimal rule (decision boundary) for allocating an observation x to a particular class. Calculate the probability of correct classification for each class. Calculate the total error rate for this choice of decision boundary.

Solution.

$$p(\omega_i|x) = \frac{p(x|\omega_i)p(\omega_i)}{\sum_i^c p(x|\omega_i)p(\omega_i)} = \frac{e^{-\lambda_i}\lambda_i^x}{e^{-\lambda_1}\lambda_1^x + e^{-\lambda_2}\lambda_2^x}$$

Decision boundary. When to guess ω_1?

$$p(\omega_1|x) > p(\omega_2|x)$$
$$e^{-\lambda_1}\lambda_1^x > e^{-\lambda_2}\lambda_2^x$$

Take the log of both sides
$$-\lambda_1 + x log(\lambda_1) > -\lambda_2 + x log(\lambda_2)$$
$$x > \frac{\lambda_1 - \lambda_2}{log(\lambda_1/\lambda_2)}$$
$$x > \frac{10 - 15}{log(10/15)}$$
$$x > 12.33$$

Probability of correct classification

$$p(\omega_1 \text{ correct}) = \sum_0^{\theta} p(x|\omega_1) = \sum_0^{12} p(x|\omega_1) = 0.791$$
$$p(\omega_2 \text{ correct}) = \sum_{\theta}^{\infty} p(x|\omega_2) = \sum_{13}^{\infty} p(x|\omega_2) = 0.732$$

Total error rate

$$p(\text{error}) = 1 - p(\omega_1 \text{ correct})p(\omega_1) - p(\omega_2 \text{ correct})p(\omega_2)$$
$$p(\text{error}) = 1 - 0.791 * 0.5 - 0.732 * 0.5$$
$$p(\text{error}) = 0.238$$

$\square$

(b) Suppose instead of one, we can obtain two independent measurements x_1 and x_2 for the object to be classified. How do the allocation rules and error rates change? Calculate the revised probability of correct classification for each class. Calculate the new total error in this case.

Solution. First, let us calculate the decision boundary. We incorporate the two measurements by adding them (and implicitly averaging them). Since we have two independent measurements, averaging them will give us a better estimation of the true value of X. If we were to get a large number of measurements, the average of the measurements would approach the true measurement of X, according to the Law of Large Numbers. Because samples are i.i.d, we can use the property of independent Poisson distributions that

$$X_1 + X_2 \sim \text{Poisson}(\lambda_{x_1} + \lambda_{x_2}) = \text{Poisson}(2\lambda)(\text{since } \lambda_{x_1} = \lambda_{x_2})$$

Let us define the random variable Y, such that $Y = X_1 + X_2$.

$$p(\omega_1|y) > p(\omega_2|y)$$
$$e^{-2\lambda_1}\lambda_1^y > e^{-2\lambda_2}\lambda_2^y$$

Take the log of both sides

$$-2\lambda_1 + y log(\lambda_1) > -2\lambda_2 + y log(\lambda_2)$$
$$\frac{y}{2} > \frac{\lambda_1 - \lambda_2}{log(\lambda_1/\lambda_2)}$$
$$\frac{y}{2} > \frac{10 - 15}{log(10/15)}$$
$$\frac{(x_1 + x_2)}{2} > 12.33$$

Probability of correct classification

$$p(\omega_1 \text{ correct}) = \sum_{y=0}^{\theta} p(y|\omega_1) \qquad = \sum_{y=0}^{24} (e^{-20} * 20^y)/y! = 0.843$$
$$p(\omega_2 \text{ correct}) = \sum_{y=\theta}^{\infty} p(y|\omega_2) \qquad = \sum_{y=25}^{\infty} (e^{-30} * 30^y)/y! = 0.843$$

Total error rate

$$p(\text{error}) = 1 - p(\omega_1 \text{ correct})p(\omega_1) - p(\omega_2 \text{ correct})p(\omega_2)$$
$$p(\text{error}) = 1 - 0.843 * 0.5 - 0.843 * 0.5$$
$$p(\text{error}) = 0.157$$

□