Problem 1:

To determine whether the given functions are convex, we assess the second derivative or apply known properties about convex functions.

(a) $f(x) = e^x - 1$

The exponential function e^x is a well-known convex function since its second derivative is always positive:

First derivative: $f'(x) = e^x$

Second derivative: $f''(x) = e^x$

Since $e^x > 0$ for all x, $f(x) = e^x - 1$ is convex.

Conclusion: $f(x) = e^x - 1$ is convex.

(b) $f(x) = \sum_{i=1}^{n} \alpha_i x_{[i]}$

Given the hint, $g(x) = \sum_{j=1}^{k} x_{[j]}$ is convex. The function $f(x)$ is a weighted sum of the largest components of x, with non-increasing non-negative weights α_i.

A weighted sum of convex functions with non-negative weights is also convex. Thus, $f(x)$ is convex.

Conclusion: $f(x) = \sum_{i=1}^{n} \alpha_i x_{[i]}$ is convex.

© $f(x) = -\log(-\log(\sum_{i=1}^{m} e^{a_i x + b_i}))$

The hint states that $\log(\sum_{i=1}^{n} e^{y_i})$ is convex. This implies that the expression $\log(\sum_{i=1}^{m} e^{a_i x + b_i})$ is convex because the sum of exponentials is also a log-sum-exp function, which is convex.

The function $-\log(x)$ is decreasing and convex for $0 < x < 1$.

Therefore, the composition of a convex function $-\log(x)$ applied to $-\log(\sum_{i=1}^{m} e^{a_i x + b_i})$, where the argument is less than 1, preserves convexity.

Conclusion: $f(x) = -\log(-\log(\sum_{i=1}^{m} e^{a_i x + b_i}))$ is convex within the domain specified by $\sum_{i=1}^{m} e^{a_i x + b_i} < 1$.

Problem 2:

To find the conjugate function $f^*(y)$ of the convex function $f(x) = \frac{1}{2}x^2$, we need to use the definition of the Legendre-Fenchel transform:

$$f^*(y) = \sup_x (yx - f(x))$$

Given $f(x) = \frac{1}{2}x^2$, we have:

$$f^*(y) = \sup_x \left(yx - \frac{1}{2}x^2 \right)$$

To find this supremum, differentiate with respect to x and set the derivative to zero:

The function to maximize is $g(x) = yx - \frac{1}{2}x^2$.

Differentiate: $g'(x) = y - x$.

Set the derivative to zero: $y - x = 0$.

Thus, $x = y$.

Substitute $x = y$ back into the expression for $g(x)$:

$$f^*(y) = y(y) - \frac{1}{2}(y)^2 = y^2 - \frac{1}{2}y^2 = \frac{1}{2}y^2$$

Therefore, the conjugate function is:

$$f^*(y) = \frac{1}{2}y^2$$

Problem 3:

From equation 2:

$$y = \frac{3}{1-\lambda}$$

Insert these into the constraint $x^2 + y^2 = 16$:

$$\left(\frac{2}{1-\lambda}\right)^2 + \left(\frac{3}{1-\lambda}\right)^2 = 16$$

Simplify:

$$\frac{4}{(1-\lambda)^2} + \frac{9}{(1-\lambda)^2} = 16$$

$$\frac{13}{(1-\lambda)^2} = 16$$

$$(1-\lambda)^2 = \frac{13}{16}$$

$$1-\lambda = \pm\frac{\sqrt{13}}{4}$$

Solving for λ:

$\lambda = 1 - \frac{\sqrt{13}}{4}$

$\lambda = 1 + \frac{\sqrt{13}}{4}$

Calculate x and y for each λ:

For $\lambda = 1 - \frac{\sqrt{13}}{4}$:

$$x = \frac{2}{\frac{\sqrt{13}}{4}} = \frac{8}{\sqrt{13}}, \quad y = \frac{3}{\frac{\sqrt{13}}{4}} = \frac{12}{\sqrt{13}}$$

For $\lambda = 1 + \frac{\sqrt{13}}{4}$:

$$x = \frac{2}{-\frac{\sqrt{13}}{4}} = -\frac{8}{\sqrt{13}}, \quad y = \frac{3}{-\frac{\sqrt{13}}{4}} = -\frac{12}{\sqrt{13}}$$

To minimize the function $f(x, y) = (x-2)^2 + (y-3)^2$ subject to the constraint $g(x, y) = x^2 + y^2 - 16 = 0$, we use the method of Lagrange multipliers.

First, set up the Lagrangian:

$$\mathcal{L}(x, y, \lambda) = (x-2)^2 + (y-3)^2 + \lambda(x^2 + y^2 - 16)$$

Compute the gradients:

$$\nabla f = \begin{bmatrix} 2(x-2) \\ 2(y-3) \end{bmatrix}$$

$$\nabla g = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$$

Set the gradients equal, scaled by λ:

$$\nabla f = \lambda \nabla g$$

This gives the system of equations:

$2(x-2) = \lambda \cdot 2x$

$2(y-3) = \lambda \cdot 2y$

$x^2 + y^2 = 16$

Simplifying the equations:

$x - 2 = \lambda x \Rightarrow x(1-\lambda) = 2$

$y - 3 = \lambda y \Rightarrow y(1-\lambda) = 3$

From equation 1, solve for λ:

$$x = \frac{2}{1-\lambda}$$

From equation 2:

$$y = \frac{3}{1-\lambda}$$

Evaluate $f(x, y)$ at each critical point:

$f\left(\frac{8}{\sqrt{13}}, \frac{12}{\sqrt{13}}\right)$

$f\left(-\frac{8}{\sqrt{13}}, -\frac{12}{\sqrt{13}}\right)$

Choose the minimum value.

The minimum occurs at $\left(\frac{8}{\sqrt{13}}, \frac{12}{\sqrt{13}}\right)$, giving the minimum value of the objective function.

Problem 4:

To determine whether the function $f(x, y) = x^3 + xy + y^3 - 3x - 6y$ is convex, concave, or neither, we'll use the Hessian matrix.

Step 1: Calculate the First Derivatives

The partial derivatives are:

$$f_x = 3x^2 + y - 3$$
$$f_y = x + 3y^2 - 6$$

Step 2: Calculate the Second Derivatives

The second partial derivatives are:

$$f_{xx} = 6x$$
$$f_{yy} = 6y$$
$$f_{xy} = f_{yx} = 1$$

Step 3: Form the Hessian Matrix

The Hessian matrix H is:

$$H = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} = \begin{bmatrix} 6x & 1 \\ 1 & 6y \end{bmatrix}$$

Step 4: Analyze the Hessian Matrix

The determinant of the Hessian matrix is:

$$\det(H) = (6x)(6y) - (1)(1) = 36xy - 1$$

For a function to be convex, the Hessian matrix should be positive semidefinite ($\det(H) \geq 0$ and $f_{xx} \geq 0$) for all x and y.

For a function to be concave, the Hessian matrix should be negative semidefinite ($\det(H) \leq 0$ and $f_{xx} \leq 0$) for all x and y.

Step 5: Check Sign of Determinant and Entries

$\det(H) = 36xy - 1$ is not always positive or always negative, thus the function is neither convex nor concave in general.

Step 6: Find Points Where the Function is Not Convex

To find points where the function is not convex, set $\det(H) < 0$:

$$36xy < 1 \quad \Rightarrow \quad xy < \frac{1}{36}$$

Any point (x, y) satisfying $xy < \frac{1}{36}$ is where the function is neither convex nor concave.

For instance, the point $(0, 0)$ yields:

$$36 \cdot 0 \cdot 0 = 0 < 1$$

Thus, the function $f(x, y)$ is neither convex nor concave at $(0, 0)$.

Problem 5:

To derive the dual norm of the l_2-norm using the Lagrangian multiplier method, we start by setting the problem as follows:

Maximize $\mathbf{y}^T\mathbf{x}$ subject to the constraint $\mathbf{x}^T\mathbf{x} \leq 1$.

Step 1: Set Up the Lagrangian

The Lagrangian $\mathcal{L}$ is given by:

$$\mathcal{L}(\mathbf{x}, \lambda) = \mathbf{y}^T\mathbf{x} + \lambda(1 - \mathbf{x}^T\mathbf{x})$$

Step 2: Find the Stationary Points

To find the stationary points, take the gradient of the Lagrangian and set it to zero:

$$\nabla_{\mathbf{x}}\mathcal{L} = \mathbf{y} - 2\lambda\mathbf{x} = 0$$

Solving for $\mathbf{x}$, we get:

$$\mathbf{x} = \frac{\mathbf{y}}{2\lambda}$$

Step 3: Use the Constraint

Substitute $\mathbf{x}$ back into the constraint:

$$\left(\frac{\mathbf{y}}{2\lambda}\right)^T\left(\frac{\mathbf{y}}{2\lambda}\right) \leq 1$$
$$\frac{\mathbf{y}^T\mathbf{y}}{4\lambda^2} \leq 1$$
$$\mathbf{y}^T\mathbf{y} \leq 4\lambda^2$$

Thus, $\lambda \geq \frac{\|\mathbf{y}\|}{2}$.

Step 4: Maximize the Lagrangian

Now substitute back into the Lagrangian:

$$\mathcal{L} = \mathbf{y}^T\left(\frac{\mathbf{y}}{2\lambda}\right) + \lambda\left(1 - \left(\frac{\mathbf{y}^T\mathbf{y}}{4\lambda^2}\right)\right)$$
$$= \frac{\mathbf{y}^T\mathbf{y}}{2\lambda} + \lambda - \frac{\mathbf{y}^T\mathbf{y}}{4\lambda}$$
$$= \lambda$$

To maximize, choose $\lambda = \frac{\|\mathbf{y}\|}{2}$:

$$\mathcal{L} = \frac{\|\mathbf{y}\|}{2}$$

Conclusion

The dual norm of the l_2-norm is:

$$\langle \mathbf{y} \rangle_* = \|\mathbf{y}\|$$

Thus, we verified that the dual norm of the l_2-norm (Euclidean norm) is itself the l_2-norm.

Problem 6:

To obtain the maximum likelihood estimator (MLE) of α, we follow these steps.

(a) Derivation of the MLE for α

The probability density function is given by:

$$f(x, \alpha) = \alpha^{-2} x e^{-\frac{x}{\alpha}}$$

For a sample $(x_1, x_2, \ldots, x_n)$, the likelihood function $L(\alpha)$ is:

$$L(\alpha) = \prod_{i=1}^{n} \alpha^{-2} x_i e^{-\frac{x_i}{\alpha}}$$

Taking the natural logarithm, the log-likelihood function $\ell(\alpha)$ becomes:

$$\ell(\alpha) = \sum_{i=1}^{n}\left(-2\log(\alpha) + \log(x_i) - \frac{x_i}{\alpha}\right)$$
$$= -2n\log(\alpha) + \sum_{i=1}^{n}\log(x_i) - \frac{1}{\alpha}\sum_{i=1}^{n} x_i$$

Differentiate $\ell(\alpha)$ with respect to α and set it to zero to find the critical points:

$$\frac{d\ell}{d\alpha} = -\frac{2n}{\alpha} + \frac{1}{\alpha^2}\sum_{i=1}^{n} x_i = 0$$

Solving for α, we obtain:

$$2n\alpha = \sum_{i=1}^{n} x_i$$
$$\hat{\alpha} = \frac{1}{2n}\sum_{i=1}^{n} x_i$$

(b) Calculation with the Given Dataset

Given the dataset: $x_1 = 0.25, x_2 = 0.75, x_3 = 1.5, x_4 = 2.5, x_5 = 2$, we have $n = 5$.

Calculate $\sum_{i=1}^{5} x_i$:

$$\sum_{i=1}^{5} x_i = 0.25 + 0.75 + 1.5 + 2.5 + 2 = 7$$

Using the estimator $\hat{\alpha}$:

$$\hat{\alpha} = \frac{1}{2 \times 5} \times 7 = \frac{7}{10} = 0.7$$

Therefore, the maximum likelihood estimate of α using the given dataset is $\hat{\alpha} = 0.7$.

Problem 7:

To prove these results, we use properties of conditional expectation.
(a) Prove $E[X] = E_Y[E_X[X \mid Y]]$
The Law of Total Expectation states:

$$E[X] = E_Y[E[X \mid Y]]$$

This means we take the expected value of X by first conditioning on Y and then taking the expectation over Y.
(b) Prove $\mathrm{Var}[X] = E_Y[\mathrm{Var}_X[X \mid Y]] + \mathrm{Var}_Y[E_X[X \mid Y]]$
To prove this, we use the Law of Total Variance:

$$\mathrm{Var}[X] = E_Y[\mathrm{Var}[X \mid Y]] + \mathrm{Var}[E[X \mid Y]]$$

This law decomposes the variance of X into two components:
The expected value of the conditional variance $\mathrm{Var}[X \mid Y]$
The variance of the conditional expectation $E[X \mid Y]$
This identity balances the “spread” due to the variability of X around its conditional mean and the “spread” of these conditional means themselves.
These two proofs utilize fundamental properties of expectations and variances in the context of conditional distributions, confirming the assertions given in the problem statement.

Problem 8:

Let’s solve this systematically.
(a) Probability Mass Function (PMF) of S
There are four dice: one 4-sided, one 6-sided, and two 8-sided. Each die has an equal chance of being chosen:
$P(S=4) = \frac{1}{4}$
$P(S=6) = \frac{1}{4}$
$P(S=8) = \frac{2}{4} = \frac{1}{2}$
(b) Find $P(S=k \mid R=3)$ for $k=4,6,8$
Calculate $P(R=3 \mid S=k)$
$P(R=3 \mid S=4) = \frac{1}{4}$
$P(R=3 \mid S=6) = \frac{1}{6}$
$P(R=3 \mid S=8) = \frac{1}{8}$
Total Probability $P(R=3)$

$$\begin{aligned} P(R=3) &= \frac{1}{4} \times \frac{1}{4} + \frac{1}{6} \times \frac{1}{4} + \frac{1}{8} \times \frac{1}{2} \\ &= \frac{1}{16} + \frac{1}{24} + \frac{1}{16} \end{aligned}$$

Convert to a common denominator:

$$= \frac{3}{48} + \frac{2}{48} + \frac{3}{48} = \frac{8}{48} = \frac{1}{6}$$

Compute $P(S=k \mid R=3)$
$P(S=4 \mid R=3) = \frac{\frac{1}{4} \times \frac{1}{4}}{\frac{1}{6}} = \frac{3}{8}$
$P(S=6 \mid R=3) = \frac{\frac{1}{6} \times \frac{1}{4}}{\frac{1}{6}} = \frac{1}{4}$
$P(S=8 \mid R=3) = \frac{\frac{1}{8} \times \frac{1}{2}}{\frac{1}{6}} = \frac{3}{8}$

The 4-sided and 8-sided dice are equally likely.

© Most Likely Die if $R = 6$

Calculate $P(R = 6 \mid S = k)$

$P(R = 6 \mid S = 4) = 0$

$P(R = 6 \mid S = 6) = \frac{1}{6}$

$P(R = 6 \mid S = 8) = \frac{1}{8}$

Total Probability $P(R = 6)$

$$P(R = 6) = 0 \cdot \frac{1}{4} + \frac{1}{6} \cdot \frac{1}{4} + \frac{1}{8} \cdot \frac{1}{2}$$
$$= \frac{1}{24} + \frac{1}{16}$$

Convert to a common denominator:

$$= \frac{2}{48} + \frac{3}{48} = \frac{5}{48}$$

Compute $P(S = k \mid R = 6)$

$P(S = 6 \mid R = 6) = \frac{\frac{1}{6} \times \frac{1}{4}}{\frac{5}{48}} = \frac{4}{5}$

$P(S = 8 \mid R = 6) = \frac{\frac{1}{8} \times \frac{1}{2}}{\frac{5}{48}} = \frac{1}{5}$

When $R = 6$, the 6-sided die is most likely.