Fundamentals of Machine Learning: Theory

Author: Darien Nouri

Topics Covered

1. Poisson Distribution
2. Gradient Computation
3. Integration and PDF Properties

0.8 Question 8

Suppose,

$$f(x,y) = \frac{\sqrt{x+y}+\sqrt{x-y}}{\sqrt{x+y}-\sqrt{x-y}} + \sqrt{\sqrt{x+y}-\sqrt{x-y}}$$

What is the value of the expression

$$2y\frac{\partial^2 f}{\partial x^2} + 4x\frac{\partial^2 f}{\partial x \partial y} + 2y\frac{\partial^2 f}{\partial y^2} + 2\frac{\partial f}{\partial y}$$

at the point where $x = 5$ and $y = 4$?

Answer:

To find the value of the given expression we need to compute the first and second partial derivatives and evaluate at point $(x, y) = (5, 4)$

$$\left((x+y)^{\frac{1}{2}} - (x-y)^{\frac{1}{2}}\right)^{\frac{1}{2}} - \frac{(x-y)^{\frac{1}{2}} + (x+y)^{\frac{1}{2}}}{(x-y)^{\frac{1}{2}} - (x+y)^{\frac{1}{2}}}$$

$$\begin{aligned} f(x,y) &= \frac{\sqrt{x+y}+\sqrt{x-y}}{\sqrt{x+y}-\sqrt{x-y}} + \sqrt{\sqrt{x+y}-\sqrt{x-y}} \\ &= \frac{(\sqrt{x+y}+\sqrt{x-y})^2}{2y} + \sqrt{\sqrt{x+y}-\sqrt{x-y}} \\ &= \frac{\sqrt{x}}{y} + \sqrt{2y} \end{aligned}$$

$$\begin{aligned} \frac{\partial f}{\partial x} &= \frac{(\sqrt{x-y}+\sqrt{x+y})\left(\frac{1}{2\sqrt{x-y}} - \frac{1}{2\sqrt{x+y}}\right)}{\left(\sqrt{x-y}-\sqrt{x+y}\right)^2} - \frac{\frac{1}{2\sqrt{x-y}} - \frac{1}{2\sqrt{x+y}}}{2\sqrt{\sqrt{x+y}-\sqrt{x-y}}} - \frac{\frac{1}{2\sqrt{x-y}} + \frac{1}{2\sqrt{x+y}}}{\sqrt{x-y}-\sqrt{x+y}} \\ \frac{\partial^2 f}{\partial x^2} &= \text{For the sake of simplicity the remaing multi-page partial derivatives} \\ &\quad \text{are omitted from this latex representation} \\ \frac{\partial f}{\partial y} &= \dots \\ \frac{\partial^2 f}{\partial x \partial y} &= \dots \\ \frac{\partial^2 f}{\partial y^2} &= \dots \end{aligned}$$

0.7 Question 7

Let $X_n = f(W_n, X_{n-1})$ for $n = 1, ..., P$, for some function $f()$. Let us define the value of variable E as

$$E = \|C - X_P\|^2$$

for some constant C. What is the value of the gradient $\frac{\partial E}{\partial X_0}$?

Answer:

To compute the gradient $\frac{\partial E}{\partial X_0}$, we'll use the chain rule.

Differentiating, we have:

$$\frac{\partial E}{\partial X_P} = -2(C - X_P)$$

Starting from P and working recursively to 0:

$$\frac{\partial X_P}{\partial X_{P-1}} = \frac{\partial f(W_P, X_{P-1})}{\partial X_{P-1}}$$

$$\vdots$$

$$\frac{\partial X_1}{\partial X_0} = \frac{\partial f(W_1, X_0)}{\partial X_0}$$

Using the chain rule gives:

$$\frac{\partial E}{\partial X_0} = -2(C - X_P) \times \frac{\partial f(W_P, X_{P-1})}{\partial X_{P-1}} \times \cdots \times \frac{\partial f(W_1, X_0)}{\partial X_0}$$

0.10 Question 10

Let $X = (x_1, \ldots, x_k)$ for some fixed k, be a random variable whose probability density function is defined as:

$$f(x) = \binom{n}{x_1, \ldots, x_k} p_1^{x_1} \ldots p_k^{x_k} \tag{8}$$

where

$$\binom{n}{x_1, \ldots, x_k} = \frac{n!}{x_1! \ldots x_k!} \tag{9}$$

Also $p_j \geq 0$ for all $j = \{1, \ldots, k\}$ and $\sum_{j=1}^{k} p_j = 1$. What is the value of $E(X)$ and $V(X)$?

Answer:

The expected value for each component x_j of the vector X is:

$$\begin{aligned} E(x_j) &= \sum_{x_1, \ldots, x_k} x_j \cdot f(x) \\ &= n\, p_j (1)^{n-1} \\ &= n \cdot p_j \end{aligned}$$

The variance of each diagonal component is:

$$V(x_j) = E(x_j^2) - (E(x_j))^2$$

Given that each x_j follows a binomial distribution with parameters n and p_j, the expected values are:

$$\begin{aligned} E(x_j) &= n \cdot p_j \\ E(x_j^2) &= n \cdot p_j + n(n-1)p_j^2 \end{aligned}$$

Therefore:

$$V(x_j) = n \cdot p_j (1 - p_j)$$

0.9 Question 9

For two vectors to be orthogonal their dot products must equal 0. We can show this using Algebraic techniques and substitution of the given definitions: u_1 and u_2 to show that: $u_1^\intercal u_2 = 0$.

$$\begin{aligned}
u_1^\intercal u_2 &= u_1^\intercal \left(a_2 - \frac{u_1^\intercal a_2}{u_1^\intercal u_1} \, u_1 \right) \\
&= u_1^\intercal a_2 - \frac{u_1^\intercal a_2}{u_1^\intercal u_1} \, u_1^\intercal u_1 \\
&= u_1^\intercal a_2 - u_1^\intercal a_2 \\
&= 0
\end{aligned}$$

Therefore, u_1 and u_2 are orthogonal.

0.3 Question 3

Let the function $f(x)$ be defined as:

$$f(x) = \begin{cases} 0 & \text{if } x \geq 0 \\ \frac{1}{(1+x)} & \text{if } x < 0 \end{cases} \tag{1}$$

Answer:
To answer this question, we must first consider the qualifying properties of a PDF.

- Positive throughout: $f(x) \geq 0$ for all x
- The integral over its domain is equal to 1: $\int_{-\infty}^{\infty} f(x)dx = 1$

1. Positive throughout:
   For all $x \geq 0$, $\frac{1}{x+1}$ will always be positive.
2. Integrate over domain to check if equal to 1:

$$\begin{aligned}
\int_0^\infty f(x)\,dx &= \int_0^\infty \frac{1}{1+x}\,dx && (2)\\
&= [\ln(1+x)]_0^\infty && (3)\\
&= \ln(1+\infty) - \ln(1+0) && (4)\\
&= \infty - \ln(1) && (5)\\
&= \infty && (6)
\end{aligned}$$

Conclusion: $f(x)$ is not a PDF because it does not satisfy the second property. The integral over its domain is not equal to 1. It is equal to infinity.

0.4 Question 4:

Assume that X and Y are two independent random variables and both have the same density function:

$$f(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Answer:

If X and Y are independent, then the value of $P(X+Y \leq 1)$ can be computed by integrating over their joint density function.

$$\begin{aligned} f_{X,Y}(x,y) &= f(x) \cdot f(y) && (8) \\ &= 2x \cdot 2y && (9) \\ &= 4xy && (10) \end{aligned}$$

$$\begin{aligned} P(X+Y \leq 1) &= P(Y \leq 1 - X) && (11) \\ &= \int_0^1 \int_0^{1-x} 4xy \, dy \, dx && (12) \\ &= \int_0^1 2x(1-x)^2 \, dx && (13) \\ &= 2 \int_0^1 x^3 - 2x^2 + x \, dx && (14) \\ &= 2 \left[\frac{x^4}{4} - \frac{2x^3}{3} + \frac{x^2}{2} \right]_0^1 && (15) \\ &= 2 \left(\frac{1}{4} - \frac{2}{3} + \frac{1}{2} \right) && (16) \\ &= 2 \left(\frac{1}{12} \right) && (17) \\ &= \frac{1}{6} && (18) \end{aligned}$$

0.5 Question 5

Since X is uniformally distributed its associated PDF is 1 for $x \in [0,1]$.

$$\begin{aligned} \mathbb{E}(Y) &= \int_0^1 g(x) \, f_x(x) \, dx \\ &= \int_0^1 e^x \, dx \\ &= e - 1 \end{aligned}$$

0.6 Question 6

Suppose that the number of errors per computer program has a Poisson distribution with a mean of $\lambda = 5$. We have 125 program submissions. Let $X_1, X_2, \ldots, X_{125}$ denote the number of errors in the programs. What is the value of $P(\bar{X}_n < 5.5)$?
Answer:

$$\mathbb{E}(\bar{X}) = \lambda = 5 \qquad \mathrm{Var}(\bar{X}) = \frac{\lambda}{125} = \frac{1}{25}$$

Find Z score using the Central Limit Theoreum:

$$\begin{aligned} Z &= \frac{\bar{X_n} - \mathbb{E}(\bar{X_n})}{\sigma} \\ &= \frac{5.5 - 5}{0.2} = 2.5 \end{aligned}$$

Using the standard normal table:

$$P(\bar{X_n} < 5.5) \approx 0.9938$$