CS795/895: Fundamentals of Deep Learning (Spring 2024)
Homework Assignment 1
Name: Sushmitha Halli Sudhakara

Instructions

- Replace every `todo` command with a true value. Clearly mark the assignment numbers.
- Submit your homework to Canvas by the deadline. Canvas will automatically add a time stamp on your submissions.
- Late submission policy: Homework is worth full credit if submitted before the due time, half credit during the next 48 hours, and zero credit after that.
- Collaboration policy: You *must* write up your own final solution. If you discussed with anyone, you *must* list the names of anybody you discussed with on this assignment.

Problem 1 (20 points)

Consider the matrix $\mathbf{X}$ and the vectors $\mathbf{y}$ and $\mathbf{z}$ below:

$$\mathbf{X} = \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad \mathbf{z} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

- What is the inner product of the vectors $\mathbf{y}$ and $\mathbf{z}$? (This is also sometimes called the dot product, and is sometimes written as $\mathbf{y}^\mathsf{T}\mathbf{z}$)
- What is the product $\mathbf{X}\mathbf{y}$?
- Is $\mathbf{X}$ invertible? If so, give the inverse, and if no, explain why not.
- What is the rank of $\mathbf{X}$? Explain your answer.

Solution 1

1. Inner product of the vectors $\mathbf{y}$ and $\mathbf{z}$:

   The inner product $\mathbf{y}^\mathsf{T}\mathbf{z}$ is calculated as:

   $$\mathbf{y}^\mathsf{T}\mathbf{z} = \begin{bmatrix} 1 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = 2 + 9 = 11$$

   Therefore, the inner product of the vectors $\mathbf{y}$ and $\mathbf{z}$ is 11.

2. The product $\mathbf{X}\mathbf{y}$ is calculated as follows:

   Given:

   $$\mathbf{X} = \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

   The product $\mathbf{X}\mathbf{y}$ is:

   $$\mathbf{X}\mathbf{y} = \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \times 1 + 4 \times 3 \\ 1 \times 1 + 3 \times 3 \end{bmatrix} = \begin{bmatrix} 14 \\ 10 \end{bmatrix}$$

Therefore, the product $\mathbf{Xy}$ is $\begin{bmatrix} 14 \\ 10 \end{bmatrix}$.

3. A square matrix like $\mathbf{X}$ is invertible if and only if the determinant is not equal to 0 [$\det(\mathbf{X}) \neq 0$].

   Let's calculate the determinant of $\mathbf{X}$:
$$\mathbf{X} = \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix}$$

   The determinant $\det(\mathbf{X})$ for a 2×2 matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is given by $ad - bc$.

   So, for $\mathbf{X}$:
$$\det(\mathbf{X}) = (2 \times 3) - (4 \times 1) = 6 - 4 = 2$$

   Since $\det(\mathbf{X}) = 2$, which is not equal to 0, the matrix $\mathbf{X}$ is invertible.

   To find the inverse of $\mathbf{X}$, the formula for the inverse of a 2×2 matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is:
$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

   Thus, the inverse of $\mathbf{X}$ is:
$$\mathbf{X}^{-1} = \frac{1}{2} \begin{bmatrix} 3 & -4 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 1.5 & -2 \\ -0.5 & 1 \end{bmatrix}$$

4. Rank of matrix $\mathbf{X}$:

   - To determine the rank of matrix $\mathbf{X}$, we examine the linear independence of its rows or columns. The rank of a matrix is the maximum number of linearly independent rows or columns.

   Given matrix $\mathbf{X}$:
$$\mathbf{X} = \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix}$$

   We can observe that the two rows (or columns) of $\mathbf{X}$ are not proportional to each other, which means they are linearly independent. Specifically, there is no scalar k such that $(2, 4) = k \times (1, 3)$ or vice versa. Since both rows are linearly independent, the rank of $\mathbf{X}$ is equal to the total number of its rows (or columns), which is 2. Therefore, the rank of $\mathbf{X}$ is 2.

   - The rank of a matrix $\mathbf{X}$ can also be determined by its determinant, especially for a 2×2 matrix. For such a matrix, if the determinant is non-zero, it implies that the matrix is of full rank.

   Given matrix $\mathbf{X}$:
$$\mathbf{X} = \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix}$$

   As previously calculated, the determinant of $\mathbf{X}$ is:
$$\det(\mathbf{X}) = 2 \times 3 - 4 \times 1 = 6 - 4 = 2$$

   Since $\det(\mathbf{X}) \neq 0$, it implies that $\mathbf{X}$ is of full rank. For a 2×2 matrix, full rank means the rank is 2.

   Therefore, based on the non-zero determinant, the rank of $\mathbf{X}$ is confirmed to be 2.

Problem 2 (20 points)

If $f(x_1, x_2) = x_1 \sin(x_2)e^{-x_1}$, what is the gradient $\nabla f(x)$ of f? Recall that $\nabla f(x) = \begin{pmatrix} \partial_{x_1} f \\ \partial_{x_2} f \end{pmatrix}$.

Solution 2

To find the gradient $\nabla f(\mathbf{x})$ of the function $f(x_1, x_2)$, we need to compute the partial derivatives of f with respect to x_1 and x_2.

The gradient $\nabla f(\mathbf{x})$ is a vector of these partial derivatives, given by:

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix}$$

The partial derivative of f with respect to x_1 is obtained by differentiating $f(x_1, x_2)$ while treating x_2 as a constant:

$$\frac{\partial f}{\partial x_1} = \frac{\partial}{\partial x_2}(x_1 \sin(x_2)e^{-x_1}) = -x_1 e^{-x_1} \sin(x_2) + e^{-x_1} \sin(x_2) = \sin(x_2)[e^{-x_1} - x_1 e^{-x_1}]$$

The partial derivative of f with respect to x_2 is obtained by differentiating $f(x_1, x_2)$ while treating x_1 as a constant:

$$\frac{\partial f}{\partial x_2} = \frac{\partial}{\partial x_2}(x_1 \sin(x_2)e^{-x_1}) = x_1 e^{-x_1} \cos(x_2)$$

Thus, the gradient $\nabla f(x)$ is:

$$\nabla f(x) = \begin{pmatrix} \sin(x_2)(e^{-x_1} - x_1 e^{-x_1}) \\ x_1 e^{-x_1} \cos(x_2) \end{pmatrix}$$

Problem 3 (20 points)

1. Let X be a random variable with a finite expectation $\mathbb{E}X < \infty$. Recall that the variance of a random variable is defined as $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}X)^2]$. Prove that $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.

2. What is the mean, variance, and entropy of a Bernoulli(p) random variable?

Solution 3

1. Proof for $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$:

   Let's start from $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}X)^2]$.

   By expanding the square, we get:

   $$\mathrm{Var}(X) = \mathbb{E}[X^2 - 2X\mathbb{E}X + (\mathbb{E}X)^2]$$

   Now, taking the expectation of each term separately:

   $$\mathrm{Var}(X) = \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[(\mathbb{E}X)^2]$$

   Note that $\mathbb{E}[(\mathbb{E}X)^2] = (\mathbb{E}[X])^2$ since $\mathbb{E}X$ is a constant. Thus, we have:

   $$\mathrm{Var}(X) = \mathbb{E}[X^2] - 2(\mathbb{E}[X])^2 + (\mathbb{E}[X])^2$$

Simplifying this expression gives us:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Hence proved that $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

2. The mean, variance, and entropy of a Bernoulli(p) random variable are as follows:

   1. **Mean**: The mean of a Bernoulli random variable is p. This is because a Bernoulli random variable takes the value 1 with probability p and 0 with probability $1-p$, so the mean is $1 \cdot p + 0 \cdot (1-p) = p$.

   2. **Variance**: The variance is given by $p(1-p)$. It's calculated as the expectation of the squared deviation from the mean, which in this case is $p(1-p)$ because the possible values are 0 and 1.

   3. **Entropy**: The entropy, which measures the uncertainty of the random variable, is $-p\log_2(p) - (1-p)\log_2(1-p)$. This is derived from the definition of entropy in information theory, where the entropy of a Bernoulli trial is the sum of the probabilities of each outcome multiplied by the logarithm of their reciprocal probabilities.

Problem 4 (20 points)

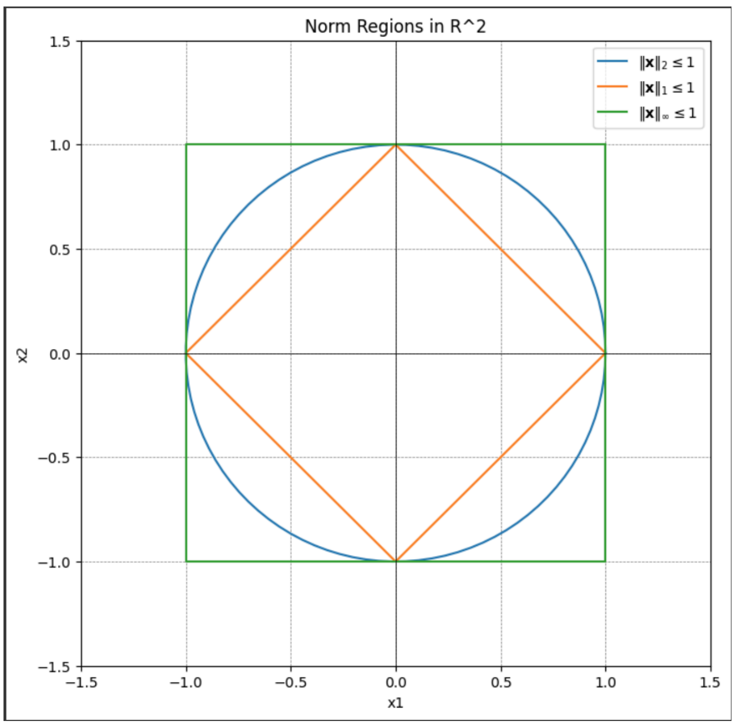
Draw the regions corresponding to vectors $\mathbf{x} \in \mathbb{R}^2$ with the following norms:

1. $\|\mathbf{x}\|_2 \leq 1$ (Recall $\|x\|_2 = \sqrt{\sum_i x_i^2}$)
2. $\|\mathbf{x}\|_1 \leq 1$ (Recall $\|x\|_1 = \sum_i |x_i|$)
3. $\|\mathbf{x}\|_\infty \leq 1$ (Recall $\|x\|_\infty = \max_i |x_i|$)

Solution 4

1. The region for $\|\mathbf{x}\|_2 \leq 1$ is a circle with radius 1 centered at the origin.
2. For $\|\mathbf{x}\|_1 \leq 1$, the region is a diamond shape with corners at (1,0), (0,1), (-1,0), and (0,-1).
3. The region where $\|\mathbf{x}\|_\infty \leq 1$ is a square with vertices at (1,1), (1,-1), (-1,1), and (-1,-1).

```python
import matplotlib.pyplot as plt
import numpy as np

# Circle for ||x||_2 <= 1
theta = np.linspace(0, 2*np.pi, 100)
x_circle = np.cos(theta)
y_circle = np.sin(theta)

# Diamond for ||x||_1 <= 1
x_diamond = [0, 1, 0, -1, 0]
y_diamond = [1, 0, -1, 0, 1]

# Square for ||x||_infty <= 1
x_square = [1, 1, -1, -1, 1]
y_square = [1, -1, -1, 1, 1]

# Plotting
plt.figure(figsize=(8, 8))
plt.plot(x_circle, y_circle, label=r'$\|\mathbf{x}\|_2 \leq 1$')
plt.plot(x_diamond, y_diamond, label=r'$\|\mathbf{x}\|_1 \leq 1$')
plt.plot(x_square, y_square, label=r'$\|\mathbf{x}\|_\infty \leq 1$')

plt.xlim(-1.5, 1.5)
plt.ylim(-1.5, 1.5)
plt.gca().set_aspect('equal', adjustable='box')
plt.axhline(0, color='black',linewidth=0.5)
plt.axvline(0, color='black',linewidth=0.5)
plt.grid(color = 'gray', linestyle = '--', linewidth = 0.5)
plt.legend()
plt.title('Norm Regions in R^2')
plt.xlabel('x1')
plt.ylabel('x2')
plt.show()
```



References Cited

[1] Towards Data Science, *Vector Norms Explained*, Available at: `https://towardsdatascience.com/vector-norms-explained-e72bf26e4a38`.

[2] ResearchGate, *The geometric interpretation of different norms in R^2*, Available at: `https://www.researchgate.net/figure/The-geometric-interpretation-of-different-norms-in-R-2_fig4_367170027`.

Problem 5 (20 points)

Sampling from a distribution. Use the Python libraries `numpy` and `matplotlib`.

1. Draw 100 samples $\mathbf{x} = [x_1, x_2]$ from a 2-dimensional Gaussian distribution with mean $[0, 0]$ and identity covariance matrix, i.e., $p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d}} \exp\left(-\frac{\|\mathbf{x}\|^2}{2}\right)$. Plot them on a scatter plot (x_1 vs. x_2).
2. How does the scatter plot change if you double the variance of each component (x_1 & x_2)?

Solution 5

When you double the variance for each component in a 2D Gaussian distribution, the resulting scatter plot shows points that are spread further apart. This broader spread indicates more variation in each direction, meaning that the data points are more scattered around the mean than they are in the original distribution with lower variance.

```python
import numpy as np
import matplotlib.pyplot as plt

n_samples = 100

mean = [0, 0]

cov_identity = [[1, 0], [0, 1]]
cov_doubled = [[2, 0], [0, 2]]


samples_identity = np.random.multivariate_normal(mean, cov_identity, n_samples)
samples_doubled = np.random.multivariate_normal(mean, cov_doubled, n_samples)

fig, axes = plt.subplots(1, 2, figsize=(12, 6))

# Plotting the first scatter plot (identity covariance matrix)
axes[0].scatter(samples_identity[:, 0], samples_identity[:, 1])
axes[0].set_title('Identity Covariance Matrix')
axes[0].set_xlabel('$x_1$')
axes[0].set_ylabel('$x_2$')
axes[0].grid(True)

# Plotting the second scatter plot (doubled variance)
axes[1].scatter(samples_doubled[:, 0], samples_doubled[:, 1])
axes[1].set_title('Doubled Variance Matrix')
axes[1].set_xlabel('$x_1$')
axes[1].set_ylabel('$x_2$')
axes[1].grid(True)

plt.show()
```