$\mathbf{a_i}$ to AI 3.0

Mathematics Club × AI Club

24th October 2024

General Information

Consider a random variable X, which takes value x_1 with probability $\mathbb{P}(x_1)$, value x_2 with probability $\mathbb{P}(x_2)$, and so on. The expected value of X, $\mathbb{E}[X]$ is defined as

$$\mathbb{E}[X] = \sum_{i=1}^{n} x_i \cdot \mathbb{P}(x_i)$$

Expectation satisfies an important property called the **Linearity of Expectation**:

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

$$\mathbb{E}[cX] = c\mathbb{E}[X]$$

In general if $X_1, X_2, ..., X_N$ are random variables and $c_1, c_2, ..., c_3$ are constants then

$$\mathbb{E}\left[\sum_{i=1}^{n} c_i X_i\right] = \sum_{i=1}^{n} (c_i \mathbb{E}[X_i])$$

The sigmoid, ReLU and hyperbolic tangent functions are defined as:

$$\sigma(z) = \frac{1}{1+e^{-z}}, \quad \text{ReLU}(x) = \max(0, x) \quad \text{and} \quad \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

Assume that for the sigmoid function and the hyperbolic tangent function, the following hold:

$$\mathbb{E}[\sigma(X)] \approx \sigma(\mathbb{E}[X])$$

$$\mathbb{E}[\tanh(X)] \approx \tanh(\mathbb{E}[X])$$

§1 Crazy Sinusoids

You are given the real number $\frac{\pi}{6}$, and asked to find the sine of that number. Like any sane person, you evaluate the infinite series

$$\sin\frac{\pi}{6} = \frac{\pi}{6} - \frac{1}{3!}\left(\frac{\pi}{6}\right)^3 + \frac{1}{5!}\left(\frac{\pi}{6}\right)^5 - \frac{1}{7!}\left(\frac{\pi}{6}\right)^7 + \dots$$

and observe that the series evaluates to $\frac{1}{2}$. We are immediately impressed with your approach to this question. Using the above approach for matrices, evaluate $\sin A$ where

$$A = \begin{bmatrix} \frac{\pi}{2} & 0 & 0 \\ 0 & \frac{\pi}{3} & 0 \\ 0 & 0 & \frac{\pi}{4} \end{bmatrix}$$

§2 Expecting other Expectations

Prasanna is on a vertex of a sqaure and wants to get to the opposite (across the diagonal of the square) vertex. He can walk only on the edges of a square. Unfortunately (or fortunately) he is drunk. From any vertex he chooses one of the two edges to walk *at random*. On an average, how many edges does Prasanna walk across before reaching his destination?

§3 Symmetrices

A matrix A satisfies the equation (for real numbers a and b)

$$A^4 + aA^3 + aA + I = bA^2$$

and $\det(A + (1+i)I) = 0$. Calculate $\det(2A + (1+i)I)$.

§4 Infinite Progression

Deenabandhan is trying to escape from SGR (the mess). He is initially at the position $(2, 4)$. In each move of his escape, he moves from (x_1, y_1) to (x_2, y_2) such that:

$$x_2 = x_1 + 7y_1$$

$$y_2 = \frac{x_1}{21} - \frac{y_1}{6}$$

He moves this way forever. Find the sum of x coordinates of Deena at every step ($\sum x_i$).

§5 One Plus

Pranjal challenged Achintya to find the determinant of a matrix $[a_{ij}]$, where

$$a_{ij} = \begin{cases} ij, & \text{if } i \neq j, \\ 1 + ij, & \text{if } i = j \end{cases}$$

Achintya was able to solve the question easily using a certain concept mentioned in the Mathematics Club DC Application (You need not constrain yourself to that method to solve this question). Pranjal now challenges you to find the determinant.

§6 Watch those Angles

KK inscribes a triangle in a circle by choosing three random points on the circumference and joining them. What is the probability that the center of the circle lies inside this triangle?

§7 What is your Rank?

Let r_i denote the smallest possible rank of an $i \times i$ matrix constructed in a way defined next. Its elements are 0 along the main diagonal and its remaining elements are positive real numbers. Determine $\sum_{n=1}^{n} r_i$.

§8 Don't Multiply

Atharva is drunk and his head is spinning. Assume his head is at the origin of the 3-D coordinate space. At any point in time, he is looking at the coordinates (x_1, y_1, z_1) and his next turn is towards the coordinates (x_2, y_2, z_2). He repeats this process. The two points are related as follows:

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 & \frac{\sqrt{3}}{2} \\ 0 & 1 & 0 \\ -\frac{\sqrt{3}}{2} & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}$$

He starts by looking at an arbitrary point (x_0, y_0, z_0). Find how many points he looks at (these include initial point (x_0, y_0, z_0)) before he reaches a point he has already seen before (Atharva is so drunk he has gained the ability to rotate 360 degrees in any direction).

§9 Look Closely

There are 3 matrices X, Y, Z which satisfy $X^2 = Y^2 = Z^2$ and $YX^2 = XYZ + 2I$. Find $\det(X^{12})$.

§10 Fat Matrices

Typically a system of (real) linear equations is represented as $A\mathbf{x} = \mathbf{b}$ where $A \in \mathbb{R}^{m\times n}$, $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^m$. Matrix multiplication is defined to facilitate the representation of linear equations in this form. Consider a variable matrix $\mathbf{Y} \in \mathbb{R}^{2\times 2}$ and a solution matrix $\mathbf{S} \in \mathbb{R}^{2\times 2}$ satisfying $\mathbf{TY} = \mathbf{S}$ for some operator $\mathbf{T} \in \mathbb{R}^{2\times 2\times 2\times 2}$. Call this operator a four-fat matrix (not to be confused with a 4×4 matrix). Define 'four-fat matrix multiplication' suitably. Every element of $\mathbf{T}$ is denoted as t_{ijkl}, $1 \le i, j, k, l \le 2$. If $\mathbf{T}$ was the *identity* operator (using your definition of four-fat multiplication), which of the below elements are nonzero (unity)?

$$t_{1111}, \quad t_{1112}, \quad t_{1221}, \quad t_{1212}, \quad t_{2122}, \quad t_{2112}, \quad t_{2121}$$

§11 Noisy Data: PCA to the Rescue

Shailesh provides Nishitha with a dataset containing three input features: x_1, x_2, and x_3, where x_3 is noisy. To mitigate the impact of noise and reduce the dimensionality, Nishitha decides to apply Principal Component Analysis (PCA).

x_1	x_2	x_3	l
1	-5	0	1
14	4	15	2
6	10	15	1

Table 1: Dataset with three features x_1, x_2, x_3, and a target l.

Note: The feature matrix X is transposed in table format above.

1. Help Nishitha check if the dataset is mean-centred. If it isn't, mean-centre the data.

2. After applying PCA on the mean centered data, calculate the three principal components from the dataset.
3. What is the percentage of variance explained by each principal component? Calculate the importance of each principal component based on the explained variance.

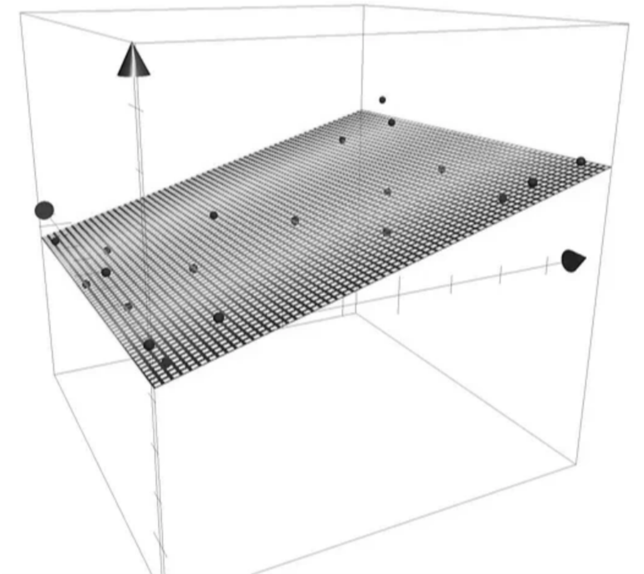
§12 Regressive Processing

After performing PCA, Nishitha shares the processed data with Harshith and Sreeram.

Harshith is working on fitting a plane to this dataset using linear regression. The equation of the plane he wants to fit is of the form

$$l = w_1 m + w_2 n + b$$

where m and n are the two important principal components obtained from Nishitha, w_1 and w_2 are the weights, and b is the bias. We randomly assign weights, biases and the learning rate as $w_1 = 0.5$, $w_2 = 0.8$, $b = 1$ and $\alpha = 0.01$. Take the loss function to be Mean Squared Error (MSE).



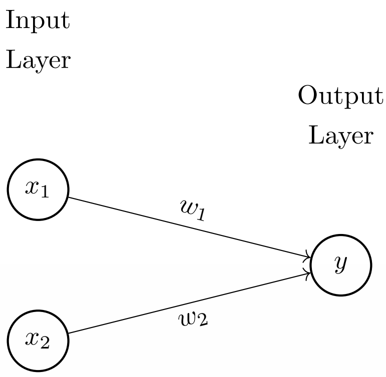
1. Perform forward propagation by calculating the predicted value $\hat{l}$ for all three data points.
2. Compute the loss.
3. Compute the gradients of the loss with respect to the parameters w_1, w_2, and b using the chain rule.
4. The gradients will be used to adjust the parameters during gradient descent. Calculate the new values for the parameters w_1, w_2, and b.

§13 Neural Network Nonsense (NNN)

Sreeram is working on building a neural network using the extracted principal components. The features used by him for the neural network are two principal components, m and n, which are the most important ones from Nishitha's PCA analysis. But Sreeram is a little mad, so he decides to pick random initial values between 0 and 10. For the weights of the neural network the probability with which he will pick a weight is given by binomial distribution and bias is set to zero. The activation for the neural network is sigmoid.

$$P(w) = \binom{10}{w} \cdot \frac{1}{2^{10}}$$

The neural network is defined as



1. What is the expected value of the output for all three sets of data points?
2. In the process of calculation of the first part, Sreeram observes something and asks Atreya what would happen if he had a more beautiful activation function like $\tanh\left(\frac{z}{2}\right)$ instead of sigmoid. Help Atreya solve this problem.

Hint: Think before you put pen to paper!

§14 Binary Bliss

You are given a binary classification dataset with two input features x_1 and x_2, and class labels $y \in \{-1, +1\}$:

x_1	x_2	y
0	1	-1
1	0	-1
1	1	$+1$
0	0	$+1$

Design a simple feed-forward neural network with the following structure: input layer containing 2 nodes, representing the two input features, one hidden layer with 2 nodes and an output layer with one node. Each node in both the hidden and output layers should have a bias term associated with it. Take ReLU as the activation function for the network.

*****

Answer Sheet

Team Name:

Name of Member 1:

Roll no. of Member 1:

Contact of Member 1:

Name of Member 2:

Roll no. of Member 2:

Contact of Member 2:

---

§1

§2

§3

§4

§5

§6

§7

§8

§9

§10

§11 1.

2.

3.

§12 1.

2.

3.

4.

§13 1.

2.

§14

Answer Key

§1 sine of values

§2 4

§3 0

§4 -91

§5 $1 + \sum n^2$

§6 0.25

§7 $3n - 4$

§8 24

§9 1

§10 $t_{1111}, t_{1212}, t_{2121}$