Maths for Machine Learning

Course Duration: 08 days (64 hours) -------> 5 days course content + 3 days exercise

Module 1: Analytic Geometry

- **Norms**
- **Inner Products**
- **Lengths and Distances**
- **Angles and Orthogonality**
- **Orthonormal Basis**
- **Orthogonal Complement**
- **Inner Product of Functions**
- **Orthogonal Projections**
- **Rotations**

Module 2: Matrix Decomposition

- **Cholesky Decomposition**
- **Singular Value Decomposition**
- **Matrix Approximation**

Module 3: Vector Calculus

- **Gradients of Vector-Valued Functions**
- **Gradients of Matrices**
- **Useful Identities for Computing Gradients**

- **Backpropagation and Automatic Differentiation**
- **Higher-Order Derivatives**
- **Linearization and Multivariate Taylor Series**

Module 4: Probability and Distribution

- **Conjugacy and the Exponential Family**
- **Change of Variables/Inverse Transform**

Module 5: Machine Learning Applications

- **Support Vector Machines**
- **Gaussian Mixture Models**
- **Principal Component Analysis (PCA)**

Practice questions for the topics done previously

Module 1: Introduction to Linear Algebra

Basic Algebra & Functions

1. Solve the equation $2x + 3 = 7$. What is the value of x?
2. Given the function $f(x) = 2x^2 - 3x + 1$, find $f(2)$.

Scalars, Vectors, and Matrices

1. Compute the magnitude of the vector $\mathbf{v} = [3, 4]$.
2. Add the vectors $\mathbf{v} = [1, 2, 3]$ and $\mathbf{w} = [4, 5, 6]$.

Matrix Operations and Properties

1. Multiply the matrices $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix}$.
2. Find the determinant of the matrix $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$.

Linear Independence and Span; Basis and Dimension

1. Determine if the vectors $\mathbf{v}_1 = [1, 2]$, $\mathbf{v}_2 = [2, 4]$ are linearly independent.
2. Find a basis for the vector space spanned by $\mathbf{v}_1 = [1, 0, 0]$, $\mathbf{v}_2 = [0, 1, 0]$, and $\mathbf{v}_3 = [0, 0, 1]$.

Eigenvalues and Eigenvectors

1. Calculate the eigenvalues of the matrix $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$.
2. Find the eigenvectors corresponding to the eigenvalues of the matrix $A = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$.

Module 2: Calculus for AI

Trigonometry

1. Find the exact value of $\sin(30^\circ)$, $\cos(45^\circ)$, and $\tan(60^\circ)$.
2. Simplify the expression $\sin^2(x) + \cos^2(x)$.

Limits, Derivatives, and Integrals

1. Evaluate the limit $\lim_{x \to 3} \frac{x^2-9}{x-3}$.
2. Find the derivative of $f(x) = 3x^3 - 5x^2 + 2x - 1$.

Chain Rule and Gradient Descent

1. Use the chain rule to differentiate the function $f(x) = (2x^2 + 3x + 1)^5$.
2. Perform one iteration of gradient descent to minimize the function $f(x) = x^2 + 4x$ with a learning rate of 0.1, starting from $x = 2$.

Multivariable Calculus and Partial Derivatives

1. Calculate the partial derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ for the function $f(x, y) = x^2y + y^3$.
2. Find the gradient vector of the function $f(x, y) = 3x^2 + 4y^2$.

Module 3: Probability and Statistics

Probability Theory

1. A fair die is rolled. What is the probability of getting a number greater than 4?
2. What is the probability of drawing a red card from a standard deck of 52 playing cards?

Random Variables and Distributions

1. For a random variable X with the probability mass function $P(X = x) = \frac{1}{2^x}$ for $x = 1, 2, 3, \ldots$, find the expected value $E(X)$.
2. If a random variable X is normally distributed with mean 0 and variance 1, what is the probability that X is between -1 and 1?

Bayesian Inference and Bayes' Rule

1. A test for a disease has a 95% sensitivity and 90% specificity. If the prevalence of the disease is 1%, what is the probability that a person who tests positive actually has the disease?
2. Compute the posterior probability given a prior probability of 0.7, a likelihood of 0.5, and a marginal likelihood of 0.6.

Hypothesis Testing and Confidence Intervals

1. Perform a hypothesis test to determine if a coin is fair, given that it lands heads 60 times out of 100 flips.
2. Construct a 95% confidence interval for the mean of a dataset with a sample mean of 50, a standard deviation of 5, and a sample size of 25.

Module 4: Mathematical Reasoning

Set Theory and Its Applications in AI

1. Given sets $A = \{1, 2, 3\}$ and $B = \{2, 3, 4\}$, find $A \cup B$, $A \cap B$, and $A \setminus B$.
2. If $A \subset B$ and $B \subset C$, prove that $A \subset C$.

Mathematical Induction and Recursion

1. Prove by induction that the sum of the first n positive integers is $\frac{n(n+1)}{2}$.
2. Write a recursive function to compute the nth Fibonacci number.

Graph Theory and Network Models

1. Draw a graph with 5 vertices and 7 edges. Determine if the graph is connected.
2. For the graph G with adjacency matrix $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, determine the degree of each vertex.

Decision Theory and Game Theory in AI

1. In a game where two players can either "Cooperate" or "Defect", the payoff matrix is given as:

	Cooperate	Defect
Cooperate	$(3, 3)$	$(0, 5)$
Defect	$(5, 0)$	$(1, 1)$

   What is the Nash equilibrium of the game?

2. Solve a decision problem where you have to choose between two options: A with a reward of 10 and probability 0.7, and B with a reward of 15 and probability 0.5.

KOENIG
step forward

Module 5: Optimization for AI

Convex Optimization

1. Show that the function $f(x) = x^2 + 3x + 4$ is convex.
2. Minimize the function $f(x) = 2x^2 + 3x + 1$ using analytical methods.

Gradient Descent and Stochastic Gradient Descent

1. Perform three iterations of gradient descent on the function $f(x) = x^2 + 6x + 9$ with a learning rate of 0.1, starting from $x = 0$.
2. Explain how stochastic gradient descent differs from batch gradient descent.

Newton's Method and Quasi-Newton Methods

1. Use Newton's method to find the root of the function $f(x) = x^3 - 2x - 5$ starting from an initial guess $x_0 = 2$.
2. Compare Newton's method with the gradient descent method in terms of convergence speed and computational cost.

Module 6: Linear Regression and Regularization

Linear Regression

1. Given the dataset $(x_1, y_1) = (1, 2)$, $(x_2, y_2) = (2, 3)$, $(x_3, y_3) = (3, 5)$, fit a linear regression model and find the equation of the best-fit line.
2. Calculate the coefficient of determination R^2 for a linear regression model where the total sum of squares (SST) is 100 and the sum of squared residuals (SSE) is 20.

Least Squares Estimation

1. Derive the normal equation for a simple linear regression model.
2. Use the least squares method to fit a line through the points $(1, 1)$, $(2, 2)$, and $(3, 3)$.

Ridge Regression and Lasso

1. Implement ridge regression on a dataset with multicollinearity and compare the results with ordinary least squares regression.
2. Explain how lasso regression can be used for feature selection.

Logistic Regression

1. Given a dataset with binary outcomes, fit a logistic regression model and interpret the coefficients.
2. Explain how the logistic regression model estimates the probability of a binary outcome.

Module 7: Neural Networks

Introduction to Neural Networks

1. Construct a simple feedforward neural network with one hidden layer of 5 neurons and an output layer with 2 neurons. Specify the activation functions for each layer.
2. Compute the output of a neural network with weights $W = \begin{bmatrix} 0.2 & 0.4 \\ 0.5 & 0.9 \end{bmatrix}$ and input $\mathbf{x} = [1, 0]$.

Backpropagation

1. Derive the backpropagation equations for a neural network with a single hidden layer.
2. Given a neural network with 2 inputs, 2 hidden neurons, and 1 output, perform one step of backpropagation to update the weights.

Training Neural Networks

1. Explain how overfitting can occur in neural networks and suggest methods to prevent it.
2. Compare the performance of a neural network trained with and without dropout regularization.

Convolutional Neural Networks (CNNs)

1. Design a CNN architecture for classifying images in the CIFAR-10 dataset. Specify the filter sizes, pooling layers, and fully connected layers.
2. Explain how pooling layers help in reducing the dimensionality of feature maps in a CNN.

Recurrent Neural Networks (RNNs)

1. Implement a simple RNN for predicting the next word in a sequence given a training dataset of sentences.
2. Compare and contrast RNNs and LSTMs in terms of handling long-term dependencies in sequential data.