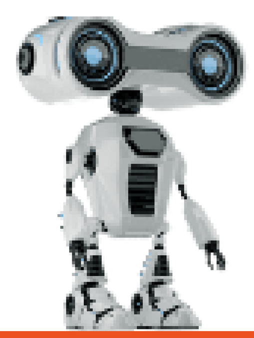
APPLIED MATHEMATICS IN MACHINE LEARNING

BASIC CONCEPTS IN ML



- Supervised Learning
- Unsupervised Learning
- Regression
- FeatureEngineering
- Ensemble Learning
- Neural Networks
- Classification
- Clustering
- EvaluationMetrics
- Dimensionality

PROBABILITY DISTRIBUTIONS

Gaussian (Normal) Distribution:

Formula: $f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

where:

- μ is the mean of the distribution.
- σ^2 is the variance of the distribution.
- x is the variable.

Bernoulli Distribution:

Formula: $f(x|p) = p^x \cdot (1-p)^{1-x}$

where:

- p is the probability of success (typically 1).
- x takes the value 1 for success and 0 for failure.

DERIVATIVES AND GRADIENTS

Derivative of a Function:

The derivative of a function $f(x)$ with respe

Gradient of a Scalar Function:

If $f : \mathbb{R}^n \to \mathbb{R}$ is a scalar function, then the partial derivatives:

$\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right)$

Partial Derivative:

If f is a function of several variables, the par variables, say x_i, is denoted by $\frac{\partial f}{\partial x_i}$.

VECTORS AND MATRICES

$V = s^3$

- Vectors represent data with both magnitude and direction.
- Matrices enable the repesentation of linear transformations, crucial in mathematics and computing.

MATHEMATICAL FOUNDATIONS OF MACHINE LEARNING

- Math & ML refers to the critical intersection where mathematical principles and techniqu converge with the field of machine learning, enabling the understanding, analysis.
- Development of algorithms to solve complex problems efficiently.

Eigen values used in Machine Leanring

Eigenvalues (λ) in machine learning are computed using the formula:

$\det(A - \lambda I) = 0$

Where:

- A is the square matrix.
- I is the identity matrix.
- λ are the eigenvalues.

Eigen vectors used in Machine Leanring!

Eigen Vectors (v) in machine learning are computed using the formula:

$(A - \lambda I)v = 0$

Where:

- A is the square matrix.
- λ is the eigenvalue.
- I is the identity matrix.

MACHINE LEARNING APPLIED IN STATISTICAL APPLICATIONS

Feature Distribution Visualization: Bar graphs can be used to visualize the distribution of features in the dataset. For categorical variables, bar graphs display the frequency or proportion of each category. This visualization helps in understanding the relative importance of different categories and can inform feature selection or preprocessing decisions.

Class Distribution in Classification Problems: In classification tasks, it's important to understand the distribution of classes in the dataset. Bar graphs can represent the class distribution, showing the frequency or proportion of each class label. Class imbalance can be identified through such visualizations, guiding the selection of appropriate evaluation metrics or sampling strategies.

STATISTICAL MEASURES :

Mean:

The mean μ of a dataset $x_1, x_2, \ldots, x_n$ is calculated as:

$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$

Variance:

The variance σ^2 measures the average squared deviation of each data point

$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$

Standard Deviation:

The standard deviation σ is the square root of the variance:

$\sigma = \sqrt{\sigma^2}$

HYPOTHESIS TESTING :

Hypothesis Testing (t-test):

For testing the difference between two population

$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

where $\bar{x}_1$ and $\bar{x}_2$ are the sample means, s_1^2 and s_2^2

LOSS FUNCTION

Loss Function: In machine learning, the loss function measures how well a model's predictions match the actual target valuesues and true values.

Mean Squared Error (MSE):

- Used primarily in regression tasks.
- Calculates the average of squared difference
- $MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

ACTIVATION FUNCTION:

Activation Function: Activation functions introduce non-linearity into neural networks, enabling them to learn complex patterns in data. They decide whether a neuron should be activated or not based on the input.

Sigmoid Function:

- Sigmoid function squashes the input values
- It is especially useful in binary classification [0, 1].
- $\sigma(x) = \frac{1}{1+e^{-x}}$

SCALARS IN MACHINE LEARNING:

- Scalars are frequently used to represent simple quantities such as error metrics, regularization parameters, or learning rates.
- For example, the loss function in a machine learning model typically outputs a scalar value representing the difference between predicted and actual values.
- Scalars are also used to represent hyperparameters, such as the number of iterations in training or the size of the batch used in stochastic gradient descent.
- Scalars are fundamental for evaluating model performance, as they quantify the quality of predictions and guide the optimization process.

VECTORS IN MACHINE LEARNING:

- Vectors are extensively used to represent features, samples, and parameters in machine learning models.
- Feature vectors are multi-dimensional arrays representing the input data points, where each element corresponds to a particular feature.
- Sample vectors represent individual data points in a dataset, with each component of the vector encoding the value of a specific feature.
- Parameter vectors are used to represent the weights or coefficients of a model, which are learned during the training process.

Scalar	Vector	Matrix	Tensor
(11)	[1,2,3]	[[1,2,3], [4,5,6], [7,8,9]]	
Shape 0	Shape 1	Shape 2	Shape n

MADE BY
HIRTHICK S
917721S011

OPTIMIZATION IN MACHINE LEARNING

Optimization in machine learning involves finding the best parameters or configurations for a model to minimize or maximize some objective function. This is typically achieved through techniques like gradient descent or its variants.

The general formula :

$$\theta_{t+1} = \theta_t - \alpha \nabla J(\theta_t)$$

θ_t represents the parameters at iteration t.

α is the learning rate, controlling the size of the steps taken during optimization.

$J(\theta)$ is the objective function or loss function being minimized with respect to θ.

$\nabla J(\theta)$ is the gradient of the objective function with respect to the parameters θ.