Introduction to Machine Learning

Agenda

1. Introduction to Machine Learning
2. Types of Machine Learning
3. Probability – An Idea
4. Linear Algebra

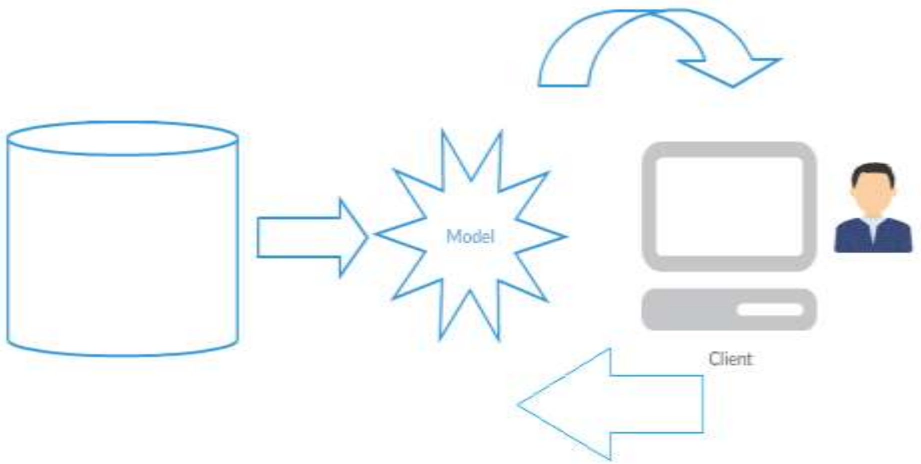
Business Problem

Retail Store:

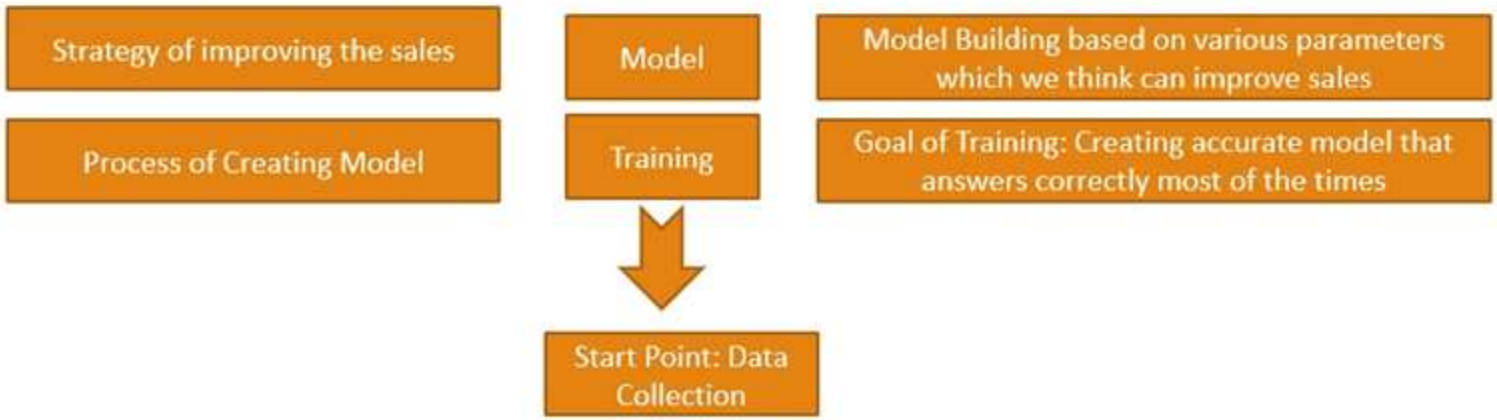
How can you improve the sales of my retail store?

Pain points and Actions

- Use unused data to get the relevant information
- Identify customers buying pattern
- Machine Learning is the solution – Build a model and deploy the same and see the magic of data science.



Process of Solving the business problem via Machine Learning



- Collect, Format, clean, reduce and rescale the data (whatever applicable and required for problem in hand and the available dataset) during data collection process

- Depending upon the type of the business problem, we choose the type of the model to be implemented. *For example:* To solve the current problem: based on various inputs taken from the customers and using the market data, let's assume, there are four probable ways of improving the sales. Now, implementation of each way involves some cost. So, we need to analyse, which parameter maximises the sales among the four ways. Also, we wants to know which is the most cost effective way to improve the sales and may be cost vs revenue ratio for each way so as to take an optimal decision for the business.

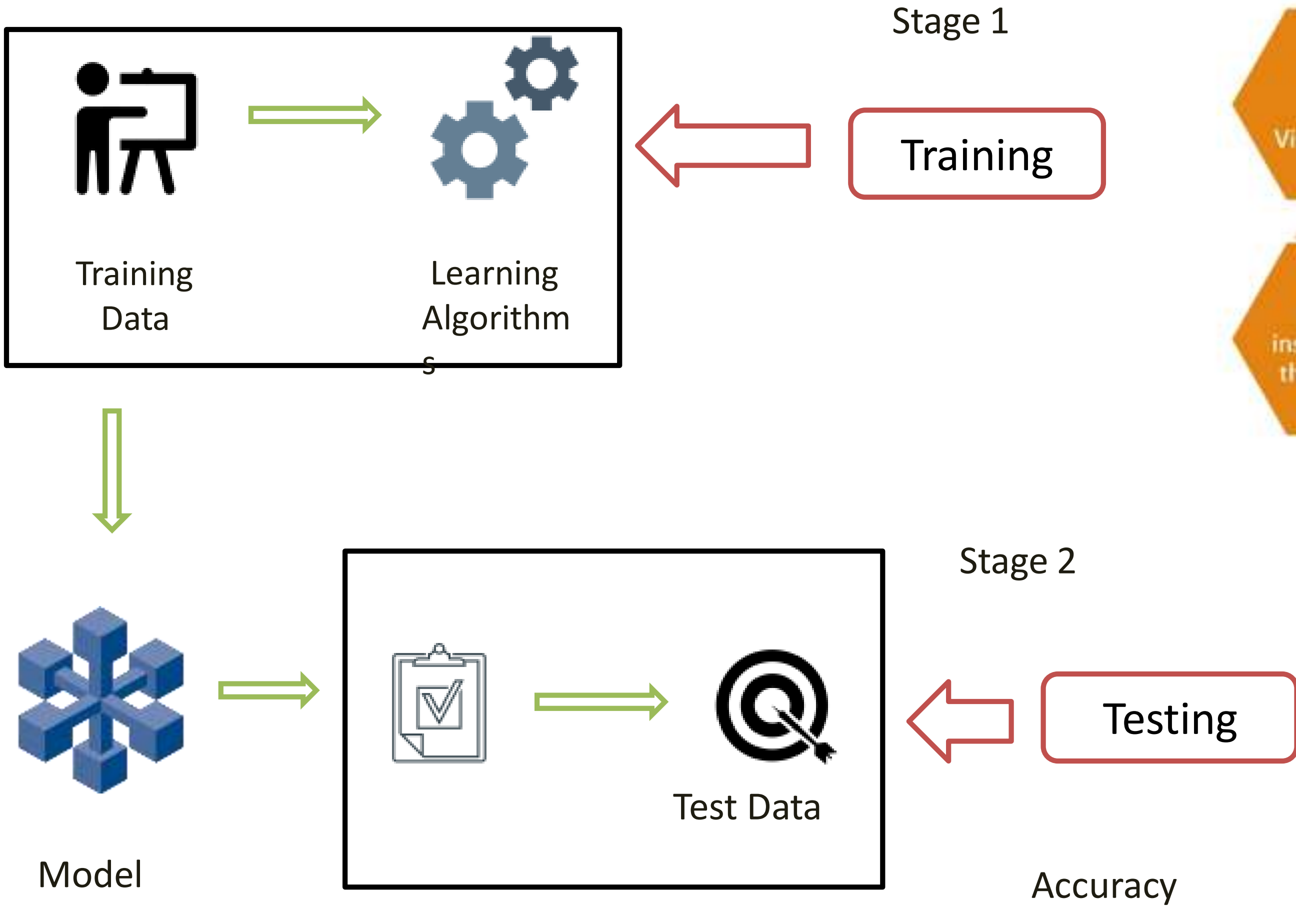
What is Machine Learning?

Machine Learning

- *Machine learning is a branch of artificial intelligence that uses statistical techniques to give machine the ability to "learn" from data, without being explicitly programmed.*

- *Helps detect patterns*
- *Get hidden insights*
- *Predict the outcome in advance which will help business to grow*

Phases of Machine Learning



Cycle of Machine Learning

Types of Machine Learning

Supervised:

It is a machine learning task where we will have to create our model using the train dataset and then validate the model using test dataset. A supervised learning algorithm analyses the training data and produces an inferred function, which can be used for mapping new datasets.

Unsupervised:

Sometimes it is very difficult to cluster our dataset and then unsupervised learning comes into picture where we don't have any data to prove the accuracy of the output.

Semi-supervised learning:

Semi-supervised learning tasks is mixture of the above two, where we use the raw data (unvalidated data) and small amount of validated data.

Reinforcement Learning:

Unlike all three mentioned above, RL doesn't have a single strategy to run. Its algorithm choses action points based on each data point.

Application of Machine Learning

Siri:

Siri is a software that adapts to user's individual preferences overtime and personalizes results.

Marketing and Sales:

Ability to capture the data for personalize shopping experience is what most of the companies are doing these days.

Application of Machine Learning

Healthcare:

Wearable devices are the best example of usage of Machine learning in healthcare domain.

Financial Services:

To prevent the data from fraud machine learning will help in the financial services domain.

Application of Machine Learning

Biometrics:

For security reasons biometrics are used a lot and is the best example of Machine Learning.

Fingerprint optical scanner:

No finger print image is created, only set of data used for the comparison.

Supervised learning algorithms

- Linear Regression
- Logistic Regression
- Decision Tree
- Random Forest
- Naïve Bayes

Used to estimate real values (Employee salary, Cost of house etc) based on continuous variable(s).

Supervised learning algorithms

Linear Regression

Used to estimate real values (Employee salary, Cost of house etc) based on continuous variable(s).

Logistic Regression

Used to estimate discrete values (True/False, Yes/No) based on given set of independent variable(s).

Decision Tree

Random Forest

Naïve Bayes

Supervised learning algorithms

Linear Regression

Used to estimate real values (Employee salary, Cost of house etc) based on continuous variable(s).

Logistic Regression

Used to estimate discrete values (True/False, Yes/No) based on given set of independent variable(s).

Decision Tree

Used for classification problems. It works for both categorical and continuous dependent variables.

Random Forest

Naïve Bayes

Supervised learning algorithms

Linear Regression	**Used to estimate real values (Employee salary, Cost of house etc) based on continuous variable(s).**
Logistic Regression	**Used to estimate discrete values (True/False, Yes/No) based on given set of independent variable(s).**
Decision Tree	**Used for classification problems. It works for both categorical and continuous dependent variables.**
Random Forest	**Random forest is an ensemble of decision trees. It gives better prediction and accuracy than decision trees.**
Naïve Bayes	

Supervised learning algorithms

Linear Regression	**Used to estimate real values (Employee salary, Cost of house etc) based on continuous variable(s).**
Logistic Regression	**Used to estimate discrete values (True/False, Yes/No) based on given set of independent variable(s).**
Decision Tree	**Used for classification problems. It works for both categorical and continuous dependent variables.**
Random Forest	**Random forest is an ensemble of decision trees. It gives better prediction and accuracy than decision trees.**
Naïve Bayes	**It is a classification technique based on Bayes theorem**

Unsupervised learning algorithms

K means

Hierarchical

Apriori

Used to segregate the data into various clusters based on the features of the dataset.

Unsupervised learning algorithms

K means

Hierarchical

Apriori

Used to segregate the data into various clusters based on the features of the dataset.

Used for creating the association rule or recommendation engine.

Libraries in Python for Data Science

Seaborn	**Focused on the visualization**
Scikit Learn	**Simple and efficient for data mining and data analysis**
Pandas	**Tool for data wrangling, designed for quick and easy Data manipulation and aggregation.**
Matplotlib	**It enables to create lot of charts such as Bar, Pie, Line etc**
NumPy	**It enables to perform operations on arrays in python**

Back to School…

Random phenomenon
Any situation/activity/game of which we know all the probable outcomes but don't know the exact outcome at given point of time.

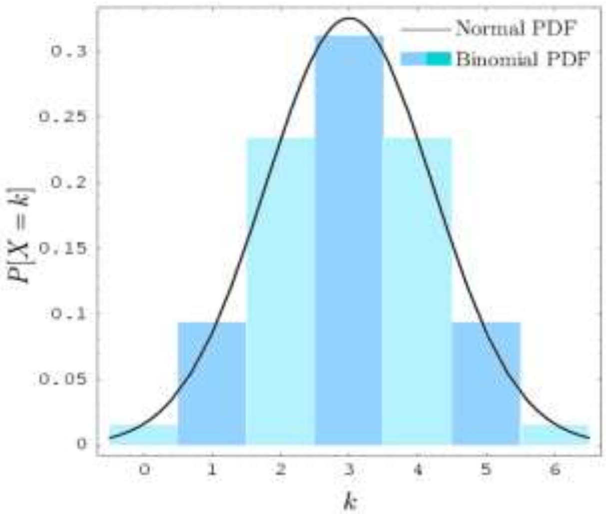
Example: Coin Toss, drawing a Card from a stack, Surveys etc.

Probability:
The outcome of a random event would be one of several possible outcomes and we can calculate the probability of the occurrence of certain outcome by analysing the random phenomena.

Probability Distribution:
The probability distribution is a description of a random phenomenon in terms of the probabilities of events.

1. Continuous
1. Discrete

Outcome	For any Random Phenomenon, each trial generates an outcome.
Event	Any set or collection of outcomes.
Sample Space	The collection of all possible outcomes is called the sample space.



Probability Distribution

Discrete Probability Distribution

- Binomial distribution.
- Geometric Distribution
- Hypergeometric distribution.
- Multinomial Distribution.
- Negative binomial distribution.
- Poisson distribution.

Continuous Probability Distribution

- Normal Distribution
- Exponential Distribution
- Logistic Distribution
- Uniform Distribution
- Multivariate Normal Distribution

Data Type

- **Discrete** → *Finite Probability e.g. Age*
- **Continuous** → *Infinite Probability e.g. Pressure*

Probability Density Function

The probability density function is the probability function which is defined for the continuous random variable. The probability density function is also called the probability distribution function or probability function. It is denoted by f (x).

Probability and its use cases

Probability is the measure of how likely something will occur
Probability outcome sums up to 1

If we roll a dice, there are six total possibilities (1,2,3,4,5,6)
Each possibility has one outcome, so each has probability of 1/6

Disjoint events – Do not have any common outcomes
The outcome of a ball cannot be six or a wicket

Non-Disjoint events – Can have common outcomes
Outcome of a ball can be a no ball and a six

Independence
Two processes are independent if the occurrence of one does not affect the probability of the other

P (A and B) = P(A) X P(B)

General Addition Rule
P(A+B)= P(A) + P(B) – P(A and B)

Note:
When A and B are disjoint, P(A and B) = 0,
hence P(A or B) = P(A) + P(B)

Linear Algebra

- A **matrix** is a rectangular **array** of numbers, symbols, or expressions, arranged in rows and columns.
- Matrices are used to describe linear equations, keep track of the coefficients of linear transformations and to record data that depend on two parameters.
- **Vectors** are simply arrays which have wrapped grow/shrink functions.
- **Linear algebra** is a sub-field of mathematics concerned with vectors, matrices, and linear transforms.
- Operation and implementation of algorithms in machine learning are been facilitated by linear algebra.
- **Linear Regression** is a method for describing the relationships between variables.
- Technique of minimizing the coefficients of the machine learning model while still being fit on the given data is called as **regularization**.
- **Principal Component Analysis** helps reducing the irrelevant columns from the given data set. This is used to create projections of high-dimensional data for training models as well as visualization.
- **One Hot Encoding** helps encoding the categorical data by forming a Boolean matrix. I.e. adding one value for the categorical value and zero in all other columns.

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

Linear Algebra

Vector in R^n is an ordered set of n real numbers.

- e.g. v = (1,6,3,4) is in R^4
- A column vector:
- A row vector:

$$\begin{pmatrix} 1 \\ 6 \\ 3 \\ 4 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 6 & 3 & 4 \end{pmatrix}$$

m-by-n matrix is an object in R^{mxn} with m rows and n columns, each entry filled with a (typically) real number:

$$\begin{pmatrix} 1 & 2 & 8 \\ 4 & 78 & 6 \\ 9 & 3 & 2 \end{pmatrix}$$

Linear Algebra

Vector norms: A norm of a vector ||x|| is informally a measure of the "length" of the vector.

$$||x||_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

- Common norms: L_1, L_2 (Euclidean)

$$||x||_1 = \sum_{i=1}^{n} |x_i| \qquad ||x||_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$$

- L_{infinity}

$$||x||_\infty = \max_i |x_i|$$

Linear Algebra

We will use lower case letters for vectors The elements are referred by x_i.

Vector dot (inner) pro

$$x^T y \in \mathbb{R} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ x_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i .$$

Vector outer product:

$$xy^T \in \mathbb{R}^{m \times n} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} = \begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & \cdots & x_2y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_my_1 & x_my_2 & \cdots & x_my_n \end{bmatrix}$$

Linear Algebra

We will use upper case letters for matrices. The elements are referred by Ai,j.

Matrix Product:

$$A \in \mathbb{R}^{m\times n} \qquad B \in \mathbb{R}^{n\times p}$$

$$C = AB \in \mathbb{R}^{m\times p}$$

$$C_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj}$$

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

$$AB = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix}$$

Special Matrices

$$\begin{pmatrix} a & 0 & 0 \\ 0 & r & 0 \\ 0 & 0 & c \end{pmatrix}$$ diagonal

$$\begin{pmatrix} a & e & c \\ 0 & r & e \\ 0 & 0 & f \end{pmatrix}$$ upper-triangular

$$\begin{pmatrix} a & d & 0 & 0 \\ c & p & r & 0 \\ 0 & f & g & h \\ 0 & 0 & l & j \end{pmatrix}$$ tri-diagonal

$$\begin{pmatrix} a & 0 & 0 \\ b & r & 0 \\ d & t & f \end{pmatrix}$$ lower-triangular

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

I (identity matrix)

Linear Algebra

Transpose: You can think of it as

- "flipping" the rows and columns

OR

- "reflecting" vector/matrix on line

e.g. $\begin{pmatrix} a \\ b \end{pmatrix}^T = \begin{pmatrix} a & b \end{pmatrix}$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^T = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

- $(A^T)^T = A$
- $(AB)^T = B^T A^T$
- $(A+B)^T = A^T + B^T$

- A set of vectors is linearly independent if none of them can be written as a linear combination of the others.
- Vectors $v_1,\ldots,v_k$ are linearly independent if $c_1v_1+\ldots+c_kv_k = 0$ implies $c_1=\ldots=c_k=0$

$$\begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

e.g. $$\begin{pmatrix} 1 & 0 \\ 2 & 4 \\ 1 & 4 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

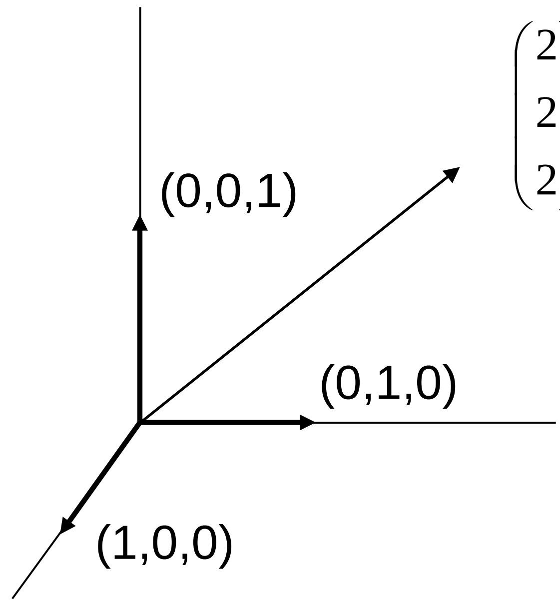
(u,v)=(0,0), i.e. the columns are linearly independent.

$$x_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad x_2 = \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix} \quad x_3 = \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}$$

x3 = −2x1 + x2

Linear Algebra

- If all vectors in a vector space may be expressed as linear combinations of a set of vectors $v_1,\ldots,v_k$, then $v_1,\ldots,v_k$ **spans** the space.
- The cardinality of this set is the **dimension** of the vector space.



$$\begin{pmatrix}2\\2\\2\end{pmatrix} = 2\begin{pmatrix}1\\0\\0\end{pmatrix} + 2\begin{pmatrix}0\\1\\0\end{pmatrix} + 2\begin{pmatrix}0\\0\\1\end{pmatrix}$$

- A **basis** is a maximal set of linearly independent vectors and a minimal set of spanning vectors of a vector space

Linear Algebra

rank(A) (the rank of a m-by-n matrix A) is

The maximal number of linearly independent columns

=The maximal number of linearly independent rows

=The dimension of col(A)

=The dimension of row(A)

If A is n by m, then

- rank(A)<= min(m,n)
- If n=rank(A), then A has full row rank
- If m=rank(A), then A has full column rank

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix}$$

Linear Algebra

Inverse of a square matrix A, denoted by A^{-1} is the *unique* matrix s.t.

- $AA^{-1} = A^{-1}A = I$ (identity matrix)

If A^{-1} and B^{-1} exist, then

- $(AB)^{-1} = B^{-1}A^{-1}$,
- $(A^T)^{-1} = (A^{-1})^T$

For orthonormal matrices

$$\mathbf{A}^{-1} = \mathbf{A}^{\mathsf{T}}$$

For diagonal matrices

$$\mathbf{D}^{-1} = \mathrm{diag}\{d_1^{-1}, \ldots, d_n^{-1}\}$$

	Scalar	Vector	Matrix
Scalar	$\frac{\mathrm{d}y}{\mathrm{d}x}$	$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}x} = \left[\frac{\partial y_i}{\partial x}\right]$	$\frac{\mathrm{d}\mathbf{Y}}{\mathrm{d}x} = \left[\frac{\partial y_{ij}}{\partial x}\right]$
Vector	$\frac{\mathrm{d}y}{\mathrm{d}\mathbf{x}} = \left[\frac{\partial y}{\partial x_j}\right]$	$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}\mathbf{x}} = \left[\frac{\partial y_i}{\partial x_j}\right]$	
Matrix	$\frac{\mathrm{d}y}{\mathrm{d}\mathbf{X}} = \left[\frac{\partial y}{\partial x_{ji}}\right]$		

$$\frac{\partial \mathbf{x}^T \mathbf{a}}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^T \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}$$

$$\frac{\partial \mathbf{a}^T \mathbf{X} \mathbf{b}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{b}^T$$

$$\frac{\partial \mathbf{a}^T \mathbf{X}^T \mathbf{b}}{\partial \mathbf{X}} = \mathbf{b}\mathbf{a}^T$$

$$\frac{\partial \mathbf{a}^T \mathbf{X} \mathbf{a}}{\partial \mathbf{X}} = \frac{\partial \mathbf{a}^T \mathbf{X}^T \mathbf{a}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{a}^T$$

$$\frac{\partial \mathbf{x}^T \mathbf{B} \mathbf{x}}{\partial \mathbf{x}} = (\mathbf{B} + \mathbf{B}^T)\mathbf{x}$$

Cognixia

A Collabera LEARNING SOLUTIONS COMPANY

THANK YOU