MACHINE LEARNING

Machine learning refers to training a system with data so that it automatically performs certain tasks and predicts the output using patterns and inferences obtained from the training dataset, without coding. It is a subset of artificial intelligence.

Primarily, in machine learning first step is that a training data set is given to a learning algorithm. Learning algorithm refers to extracting patterns from a data set and using it in appropriate places.

So this learning algorithm creates new rules based on its interpretation of the training dataset. Thus generated is known as machine learning model .Then it predicts output based on the training given to it. So more the training of a model with data, more the model learns and more accuracy will on be the result.

Applications in real life:

There are numerous applications of machine learning in real life.

Some common and important applications include data science, face sensing, preventing malicious emails, predicting rainfall ,identifying miscellaneous data etc

Principal Component Analysis:

- It is a mathematical procedure that is used mainly to reduce the number of dimensions in a data set which at the same time preserves the value (importance) of each data point.

- It comprises of an intermediate value called **Covariance** which is a parameter that shows the dependence of every attribute with every other attribute in a data set.
  (Eg, Height,Weight,BMI,etc..)

  Formula :

$$\text{Cov}(x,y) = \frac{\Sigma\,(x_i - \bar{x}) * (y_i - \bar{y})}{N}$$

Where, **Xi** , **Yi** denotes the value of that attribute which has to be summed over the whole data set. **X(bar)** and **Y(bar)** are the **mean values** of all the corresponding attribute values. **N** denotes number of data variables.

- Covariance ranges from **-1 to 1** respectively denoting **inverse relation** and **direct, closely related.**

PCA and Machine Learning:

- Machine learning is the application of Artificial Intelligence of computers to automatically learn and execute from a set of data provided without explicit programming.

- There are multiple types like **Supervised/Unsupervised machine learning algorithms.**

- PCA is a type of **Unsupervised Machine Learning Algorithm** which uses ***UN-LABELLED*** data to find a pattern and infer the data provided in various formats like plotting cluster of data points where:
    - Closely clustered points are more related
    - Sparsely located points are not much related.

Outliers:

Outliers are those points which are abnormally placed far from similar points in the plot obtained from PCA which may be caused due to experimental errors.

This is one drawback of PCA and in this case, these points should be properly handled so as to preserve accuracy of result.

PRINCIPLE COMPONENT ANALYSIS

The idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables while retaining as much as possible of the variation present in the data set.

This is achieved by transforming to a new set of variables, the *principal components (PCs)*, which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables.

MATHEMATICS BEHIND PCA

- Take the whole dataset consisting of *d+1 dimensions* and ignore the labels such that our new dataset becomes *d dimensional.*

- Compute the *mean* for every dimension of the whole dataset.
- **Compute the *covariance matrix* of the whole dataset.**

$$cov(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n} (Xi - \bar{x})(Yi - \bar{y})$$

- the result would be a *square matrix of d ×d dimensions.*

- If original matrix is like this
-

$$
\begin{array}{c}
 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5
\end{array}
\begin{array}{c}
\begin{array}{ccc} Math & English & Arts \end{array} \\
\begin{bmatrix}
90 & 60 & 90 \\
90 & 90 & 30 \\
60 & 60 & 60 \\
60 & 60 & 90 \\
30 & 30 & 30
\end{bmatrix}
\end{array}
$$

-
- Its *covariance matrix* would be
-

$$
\begin{array}{c}
 \\ Math \\ English \\ Art
\end{array}
\begin{array}{c}
\begin{array}{ccc} Math & English & Art \end{array} \\
\begin{bmatrix}
504 & 360 & 180 \\
360 & 360 & 0 \\
180 & 0 & 720
\end{bmatrix}
\end{array}
$$

-

- Shown in *Blue* along the diagonal, we see the variance of scores for each test. The art test has the biggest variance (720); and the English test, the smallest (360). So we can say that art test scores have more variability than English test scores.

- The covariance is displayed in black in the off-diagonal elements of the matrix **A**

a) The covariance between math and English is positive (360), and the covariance between math and art is positive (180). This means the scores tend to covary in a positive way. As scores on math go up, scores on art and English also tend to go up; and vice versa.

b) The covariance between English and art, however, is zero. This means there tends to be no predictable relationship between the movement of English and art scores.

- **Compute Eigenvectors and corresponding Eigenvalues**

- An Eigen vector is a vector whose direction remains unchanged when an linear transformation is applied on it.

- *Let* **A** *be a square matrix,* ***v*** *a vector and* ***λ*** *a scalar that satisfies* **A***v* ***=*** ***λ****v, then* ***λ*** *is called eigenvalue associated with eigenvector* ***v*** *of* **A***.*
- The eigenvalues of ***A*** are roots of the characteristic equation
-
- $$\det(A - \lambda I) = 0$$
- Calculating *det(A-λI)* first, *I* is an identity matrix :
-
- $$\det\left(\begin{pmatrix} 504 & 360 & 180 \\ 360 & 360 & 0 \\ 180 & 0 & 720 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right)$$
- Simplifying the matrix first, we can calculate the determinant later,
-
- $$\begin{pmatrix} 504 & 360 & 180 \\ 360 & 360 & 0 \\ 180 & 0 & 720 \end{pmatrix} - \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}$$
- $$\begin{pmatrix} 504-\lambda & 360 & 180 \\ 360 & 360-\lambda & 0 \\ 180 & 0 & 720-\lambda \end{pmatrix}$$
- Now that we have our simplified matrix, we can find the determinant of the same :
-
- $$\det\begin{pmatrix} 504-\lambda & 360 & 180 \\ 360 & 360-\lambda & 0 \\ 180 & 0 & 720-\lambda \end{pmatrix}$$
-
- $$-\lambda^3 + 1584\lambda^2 - 641520\lambda + 25660800$$
- We now have the equation and we need to solve for *λ,* so as to get the *eigenvalue of the matrix.* So, equating the above equation to zero **:**
-
- $$-\lambda^3 + 1584\lambda^2 - 641520\lambda + 25660800 = 0$$
- After solving this equation for the value of ***λ,*** we get the following value
-
- $$\lambda \approx 44.81966\ldots, \lambda \approx 629.11039\ldots, \lambda \approx 910.06995\ldots$$

- Eigenvalues
- Now, we can calculate the eigenvectors corresponding to the above eigenvalues.
- Eigen Vectors can be calculated by substituting values of lamda in matrix and solving the equation
- AX=0

Sort the eigenvectors by decreasing eigenvalues and choose k eigenvectors with the largest eigenvalues to form a *d × k dimensional* matrix W.

- the eigenvectors with the lowest eigenvalues bear the least information about the distribution of the data, and those are the ones we want to drop.

- So, *eigenvectors* corresponding to two maximum eigenvalues are :

$$\mathbf{W} = \begin{bmatrix} 1.05594 & -0.50494 \\ 0.69108 & -0.67548 \\ 1 & 1 \end{bmatrix}$$

-
- Transform the samples onto the new subspace

In the last step, we use the 2×3 dimensional matrix ***W*** that we just computed to transform our samples onto the new subspace via the equation ***y = W' × x*** where ***W'*** is the *transpose* of the matrix ***W.***

We have computed our two principal components and projected the data points onto the new subspace.