CO3 – Skilling questions (skill 9)

1. Q. Questions related to Multivariate co-variance and co-relation

Data-set

The Iris flower data set or Fisher's Iris data set is a multivariate data set introduced by the British statistician, eugenicist, and biologist Ronald Fisher in his 1936. The data set consists of 50 samples from each of three species of Iris (Iris setosa, Iris virginica and Iris versicolor). Four features were measured from each sample: the length and the width of the sepals and petals, in centimetres. Based on the combination of these four features, Fisher developed a linear dis-criminant model to distinguish the species from each other. The dataset is available in following link.

<https://www.kaggle.com/arshid/iris-flower-dataset>

Q1. A. compute the covariance between SepalLengthCm and PetalLengthCm and draw conclusion.

Q1. B. compute the covariance between SepalLengthCm and SepalWidthCm and draw conclusion.

Q1. C. compute the covariance matrix of all attributes of iris dataset with and without using the predefined method and draw conclusion.

Q2. A. compute the corelation between SepalLengthCm and PetalLengthCm and draw conclusion.

Q2. B. compute the corelation between SepalLengthCm and SepalWidthCm and draw conclusion.

Q2. C. compute the corelation matrix of all attributes of iris dataset with and without using the predefined method conclusion and draw conclusion.