* **Due** May 1 by 2:29pm

* **Points** 10

* **Submitting** a file upload

* **Available** Apr 14 at 2:30pm - May 1 at 2:29pm 17 days

This assignment was locked May 1 at 2:29pm.

Use the same [dataset](https://bits-pilani.instructure.com/courses/341/files/66861/download) comprising of 109 cricket players, filename is ‘cricketers.csv’ as used in assignment-1.

Questions:

**#1**For the K-means algorithm using two attributes namely average\_runs and bowling\_economy, which you wrote in assignment-1, extend it with different strategies (take at least two strategies) for initial choice of cluster centroid. Take K=2.

(a) Which strategy achieves better convergence (iterations count) ? Draw bar plot (X-axis has different strategies and Y-axis has iterations count).  [2 points]

(b) Is it true for other values of K = 3,4,5,6 ? On the same bar plot, draw bars with different colors for these different values of K. [2 points]

**#2** Use the elbow method to find the best value of K. Consider four attributes runs\_scored, average\_runs, wickets\_obtained, bowling\_economy.   Plot the related graph as discussed in the video lecture. [2 points]

**#3** Your task is to create a team comprising 4 batsmen, 4 bowlers, and 3 all-rounders.  Use appropriate K value. Consider two attributes runs\_scored and wickets\_obtained. Take the criteria for selecting top batsmen on the basis of runs\_scored, top bowlers on the basis of wickets\_obtained and top all-rounders on the basis of average of runs\_scored and wickets\_obtained.

(a) Output the names of cricketers who got selected in your team. [2 points]

(b) Draw a scatter plot with runs\_scored and wickets\_obtained on X-axis and Y-axis, respectively. Each selected cricketer as a data point (solid circle), use different colors for batsman, bowler and all-rounder.   [2 points]

**#4** Refer to [this](https://bits-pilani.instructure.com/courses/341/files/69240/download) notebook (GMM), perform on cricketers dataset, use runs\_scored and wickets\_obtained as the attributes. [*Not to be graded*]. This question (#4) is for your self-learning.

**Deliverables:** For questions to be graded, Q1-Q3, paste plot diagrams in a single document file and related code in single jupyter notebook.

For any query, write to me at rishabh.kaushal@wilp.bits-pilani.ac.in