Regression and classification are supervised learning approach that maps an input to an output based on example input-output pairs, while clustering is an unsupervised learning approach.

Regression과 Classification은 지도학습

Clustering은 비지도학습

**Regression**: It predicts continuous valued output. The Regression analysis is the statistical model which is used to predict the numeric data instead of labels. It can also identify the distribution trends based on the available data or historic data. Predicting a person’s income from their age, education is example of regression task.

**Classification**: It predicts discrete number of values. In classification the data is categorized under different labels according to some parameters and then the labels are predicted for the data. Classifying emails as either spam or not spam is example of classification problem.

**Clustering**: Clustering is the task of partitioning the dataset into groups, called clusters. The goal is to split up the data in such a way that points within single cluster are very similar and points in different clusters are different. It determines grouping among unlabeled data.

**'분류'**란, 소속집단의 정보를 이미 알고 있는 상태에서, 비슷한 집단으로 묶는 방법입니다.

다시말해, Label 이 있는 data를 나누는 방법으로, Supervised Learning (지도학습)의 일종입니다.

**'군집화'**란, 소속집단의 정보가 없고, 모르는 상태에서, 비슷한 집단으로 묶는 방법입니다.

즉, Label이 없는 data를 군집단위로 나누는 것으로, Unsupervised Learning (비지도학습) 의 일종입니다.



첫번째 그림에서는, 각 데이터들이 주황색/초록색/파란색이라는 색깔로 라벨링된 사전 정보가 존재합니다.   
이러한 라벨링된 데이터들의 분포에서 비슷한 집단으로 적당하게 분류하는 것이 필요한데, 이 때는 **Classification (분류)**을 사용합니다.

한번 두번째 그림에서는, 각 데이터들이 어떤 소속집단을 가지는지에 대한 사전정보가 존재하지 않습니다.  
따라서 이러한 경우에는 라벨링되지 않은 데이터들을 비슷한 집단으로 분류하는 **Clustering (군집화)** 방법이 필요합니다.