# Weekly Report 2- K-Means

Ganji Varshitha
AI20BTECH11009

---

## Introduction

K-Means is an unsupervised learning algorithm which performs partitioned clustering. Clustering helps us to understand the structure of the data by grouping it into distinct sub-groups.

## Algorithm

It is a parametric method where we need to specify the number of clusters K, which we want to divide the data into.
It is a simple and iterative algorithm.

> Initialize K random points from the dataset as centroids
>
> **repeat**
>
> > Form K clusters by assigning all points to the closest centroid
> >
> > Recompute the centroid of each cluster
>
> **until** The centroids don't change

## Key points

- The algorithm works if the dependent variables are linearly related with independent variables.
- It assumes the error terms to have constant variance and no correlation between one another.
- Since we need to compute the gradients, there is need to normalize the training data so that it does not take longer time.
- Learning rate decides how fast the algorithm converges.
- The model tends to overfit if there are many features in the dataset. Besides, there is a high chance correlation between some input variables.

## Questions

1. Bias and variance in linear regression model are _____ related.
   A. directly  B. inversely
   **Solution:**
   B. inversely

2. Explain the geometric and probabilistic interpretation of the model.
   **Solution:**
   Geometric Interpretation:
   The least-squares regression function is obtained by finding the orthogonal projection of the output vector y onto the subspace spanned by $x_1, x_2, x_3, \cdots, x_d$.
   Probabilistic Interpretation:
   Let us assume the target variable be given by the deterministic function with added gaussian noise. This gives

$$p(y|X, \mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}(y_n|\mathbf{w}^\top x_n, \beta^{-1}) \tag{1}$$

   where $\beta$ is inverse variance of zero mean Gaussian random variable. We estimate the probability model using maximum likelihood which is same as minimising the least squares error.

$$\arg\max_{\mathbf{w}} L = \arg\min_{\mathbf{w}} E \tag{2}$$

3. Do the missing values in the data affect the model?
   A. Yes  B. No
   **Solution:**
   B. No

4. Is the model generative or discriminative? Explain why.
   **Solution:**
   Linear regression is a discriminative model. As seen in the probabilistic interpretation of the model above, we learn the parameters that maximises the conditional probability P($Y|X$). This is the basic definition of discriminative model where we assume functional form of P($Y|X$) and estimate its parameters.

5. When to prefer gradient descent approach to ordinary least squares method?
   **Solution:**
   Time complexity of least squares method is $\mathcal{O}(n^3)$ whereas time complexity of gradient descent is $\mathcal{O}(n)$. Gradient descent is preferred to ordinary least squares method when n i.e number of input features is greater than 10,000.

6. Which error function is sensitive to outliers?
   A. MSE(Mean squared error)

B. MAE(Mean absolute error)

**Solution:**
A. MSE