1 Create Dataset

1. Using the below code, create **Dataset-1**.

```
1  import numpy as np
2  import random
3  #Define input array with angles from 0deg to 360deg converted to radians
4  x = np.array([i*np.pi/180 for i in range(0,360,4)])
5  #Setting seed for reproducibility
6  np.random.seed(10)
7  #Adding random noise to sine wave
8  y = np.sin(x) + np.random.normal(0,0.15,len(x))
```

2. Plot **Dataset-1** to visualize the points scattered roughly as a sine wave.

3. Create **Dataset-2** as follows:

   (a) Sample x from Gaussian Mixture Model(GMM).
   (b) Sample y from $N(W^T X, \sigma)$(Normal/Gaussian distribution with mean $(W^T X)$ and variance σ).
   (c) Plot for different values of K(x coming form K different Gaussians).

2 Implement Ridge Regression

1. For **Dataset-1**,

   (a) Find 15 powers of input array x and treat the whole as new input.
   (b) Fit the new input and output y using ridge regression with $\alpha = 0.001$.
   (c) Plot the graph showing input values and fitted curve.

2. Implement Ridge Regression for **Dataset-2** and plot graph.

3 Implement Cross Validation

Perform the following tasks for both **Dataset-1** and **Dataset-2**.

1. Find the optimum value of α using k-fold cross validation manually.

2. Find the optimum value of α using k-fold cross validation using $RidgeCV$ (Ridge regression with built-in cross-validation) which is an inbuilt function in *sklearn*.

3. Plot the graph showing input values and the curve with optimum alpha found using both methods.