In this project, I use the same dataset as project 1. It is Breast Cancer Wisconsin(Diagnostic) Data set. This data set have 30 attributes except the label and patient ID. The link to the dataset is <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29> from UCI machine learning repository.

In this project, I use scikit package which has a Gaussian Process Classifier API. So it is quite easy to implement. The first step is using getData function to import data from excel. And then use rbf kernel with r = 1 to implement GP classifier. I use a 10-fold cross validation to compute the performance. It gives me the average accuracy of 74.0% using GPC model.

Compared with linear model, GPC model performs worse than it. So for this specific dataset, I will choose linear model instead of GPC model.