ELL-409 (Machine Intelligence and

Learning)

*Assignment 2*

*Report*

*Submitted by – Mihir Gupta , 2018PH10816*

1. **Principal Component Regression**

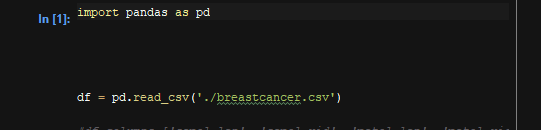
In [statistics](https://en.wikipedia.org/wiki/Statistics), principal component regression (PCR) is a [regression analysis](https://en.wikipedia.org/wiki/Regression_analysis) technique that is based on [principal component analysis](https://en.wikipedia.org/wiki/Principal_component_analysis) (PCA). More specifically, PCR is used for [estimating](https://en.wikipedia.org/wiki/Estimation) the unknown [regression coefficients](https://en.wikipedia.org/wiki/Linear_regression) in a [standard linear regression model](https://en.wikipedia.org/wiki/Linear_regression).

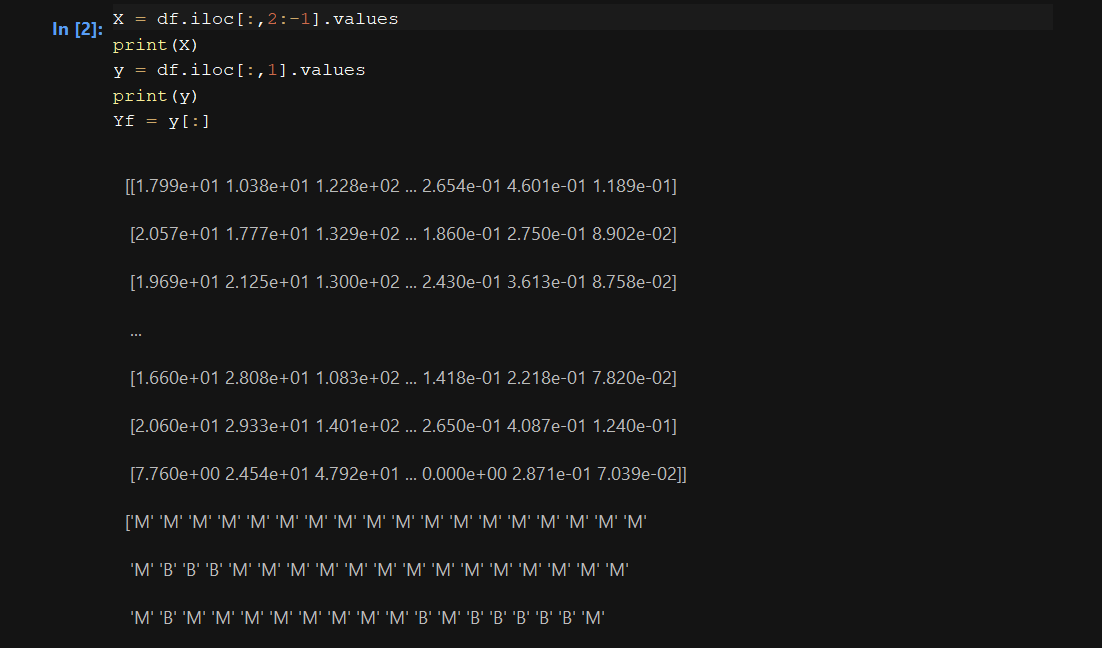
In PCR, instead of regressing the dependent variable on the explanatory variables directly, the [principal components](https://en.wikipedia.org/wiki/Principal_component_analysis) of the explanatory variables are used as [regressors](https://en.wikipedia.org/wiki/Dependent_and_independent_variables" \o "Dependent and independent variables). One typically uses only a subset of all the principal components for regression, making PCR a kind of [regularized](https://en.wikipedia.org/wiki/Regularization_(mathematics)) procedure and also a type of [shrinkage estimator](https://en.wikipedia.org/wiki/Shrinkage_estimator).

Often the principal components with higher [variances](https://en.wikipedia.org/wiki/Variance) (the ones based on [eigenvectors](https://en.wikipedia.org/wiki/Eigenvalues_and_eigenvectors) corresponding to the higher [eigenvalues](https://en.wikipedia.org/wiki/Eigenvalues_and_eigenvectors) of the [sample](https://en.wikipedia.org/wiki/Sample_mean_and_sample_covariance) [variance-covariance matrix](https://en.wikipedia.org/wiki/Covariance_matrix) of the explanatory variables) are selected as regressors. However, for the purpose of [predicting](https://en.wikipedia.org/wiki/Prediction) the outcome, the principal components with low variances may also be important, in some cases even more important.[[1]](https://en.wikipedia.org/wiki/Principal_component_regression#cite_note-1)

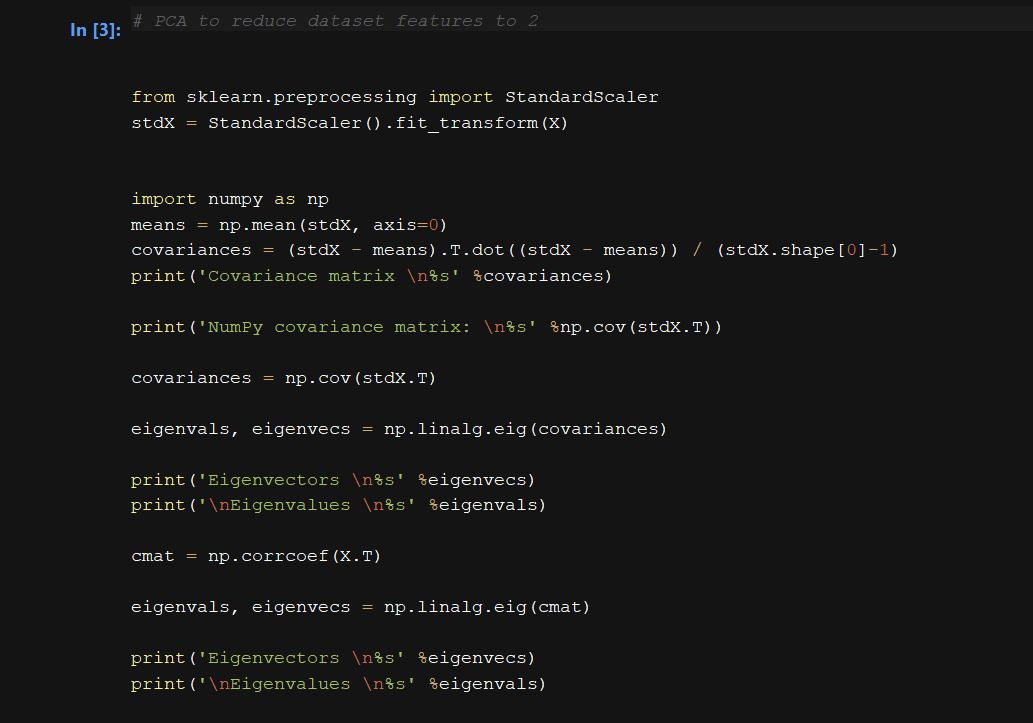
One major use of PCR lies in overcoming the [multicollinearity](https://en.wikipedia.org/wiki/Multicollinearity) problem which arises when two or more of the explanatory variables are close to being [collinear](https://en.wikipedia.org/wiki/Collinear).[[2]](https://en.wikipedia.org/wiki/Principal_component_regression#cite_note-2) PCR can aptly deal with such situations by excluding some of the low-variance principal components in the regression step. In addition, by usually regressing on only a subset of all the principal components, PCR can result in [dimension reduction](https://en.wikipedia.org/wiki/Dimensionality_reduction) through substantially lowering the effective number of parameters characterizing the underlying model. This can be particularly useful in settings with [high-dimensional covariates](https://en.wikipedia.org/wiki/High-dimensional_statistics). Also, through appropriate selection of the principal components to be used for regression, PCR can lead to efficient [prediction](https://en.wikipedia.org/wiki/Prediction) of the outcome based on the assumed model.

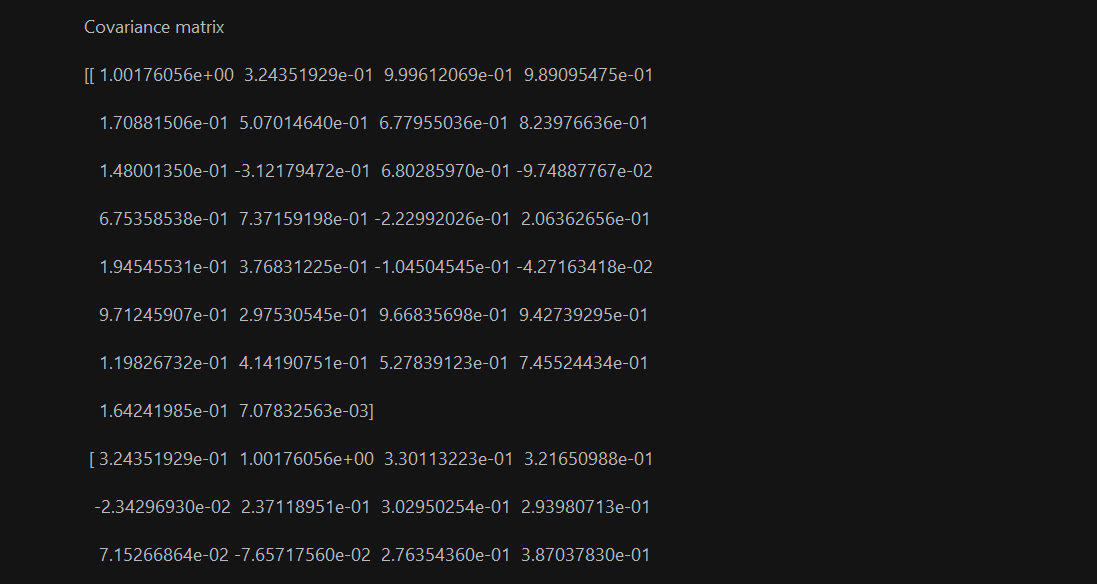
1. Dataset used – Breast cancer dataset



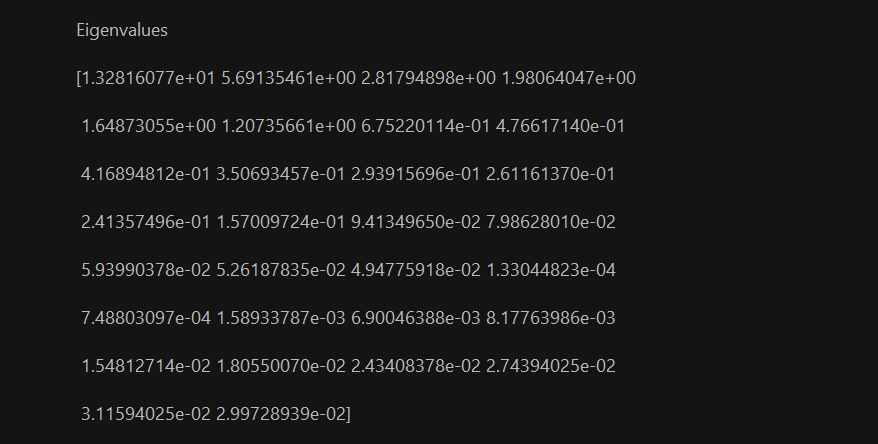


1. Generate co-variance matrix

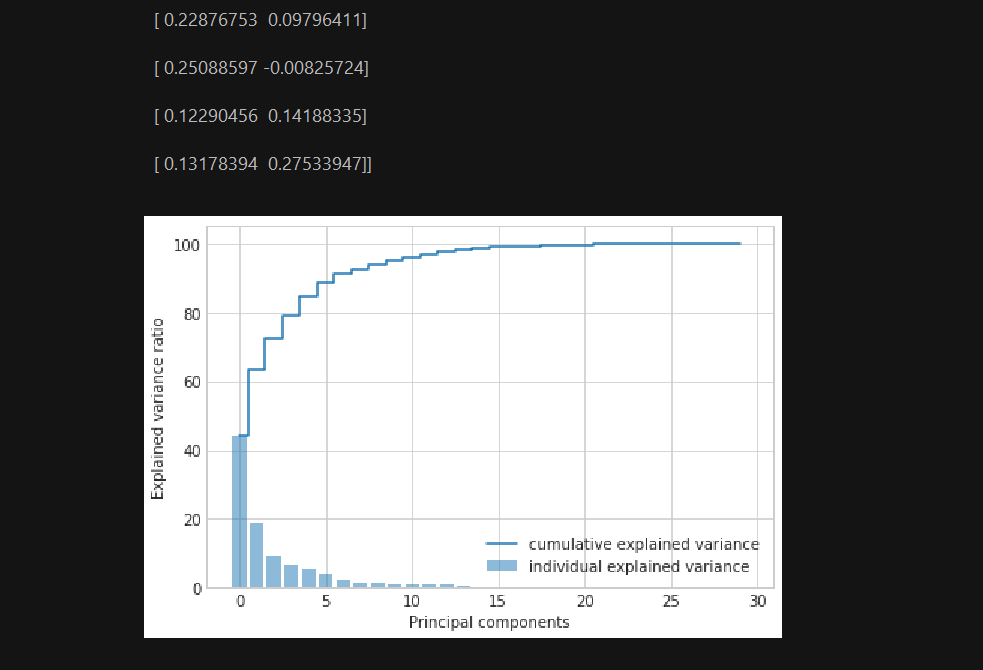




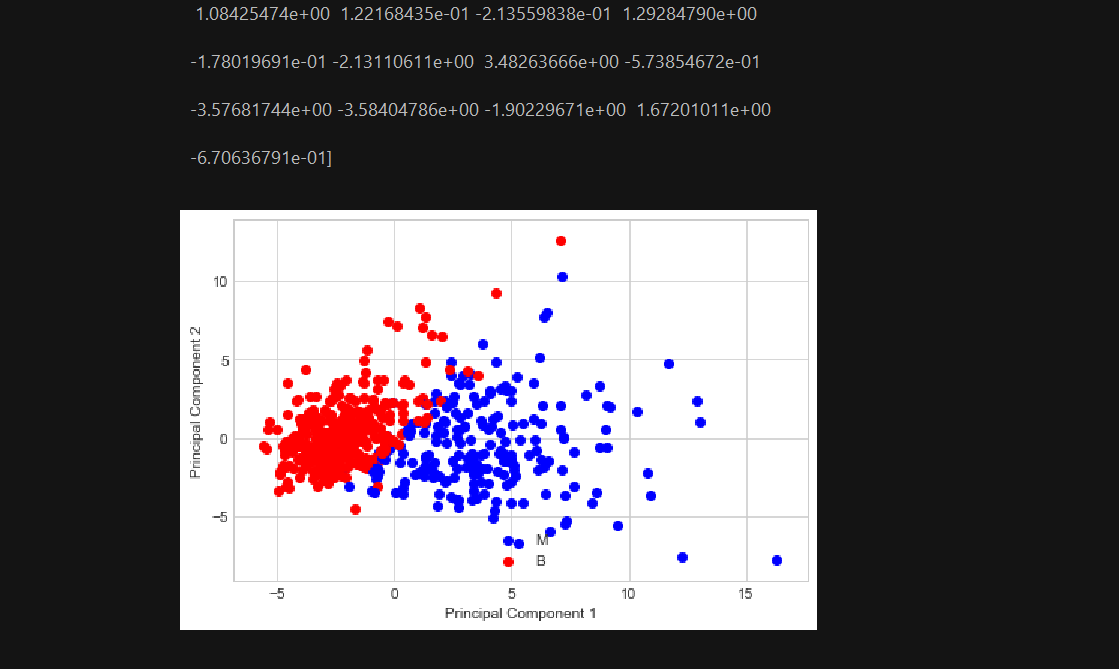
1. Determine eigenvectors and eigenvalues and print in descending order



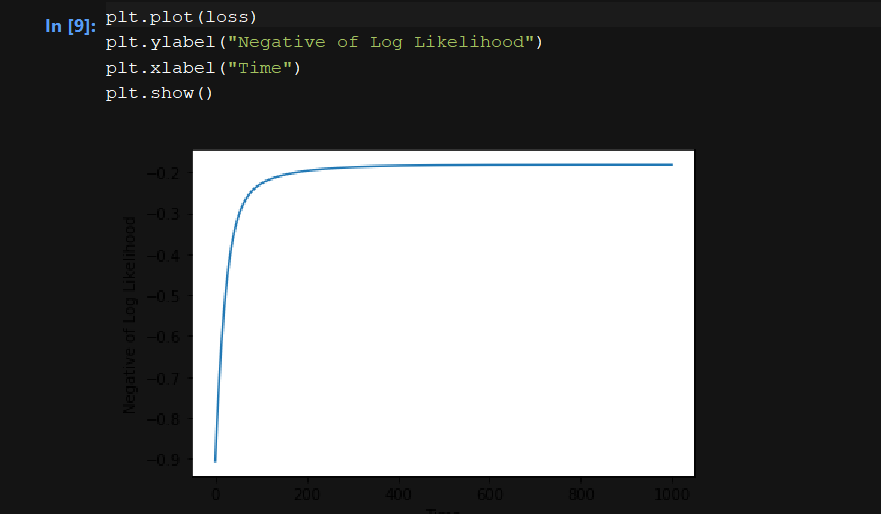
1. Determine / plot cumulative / individual variance

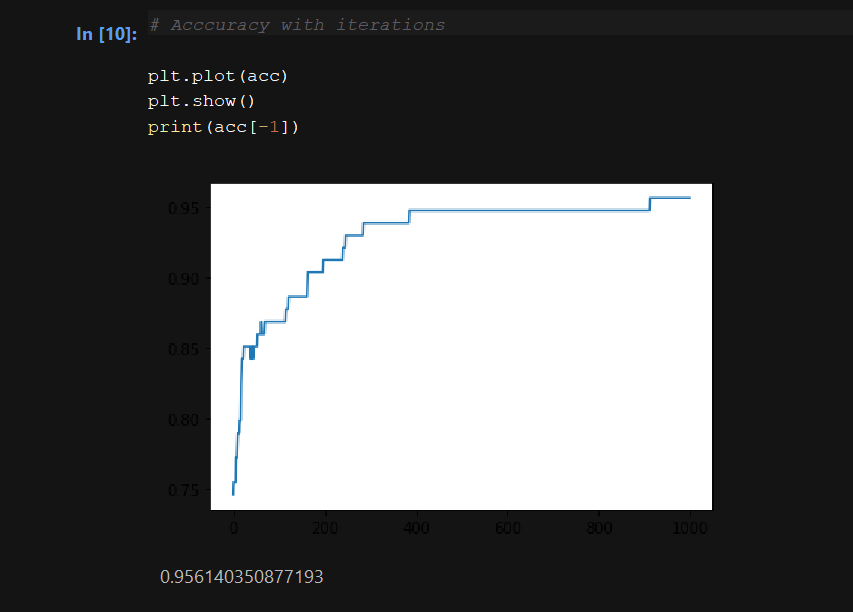


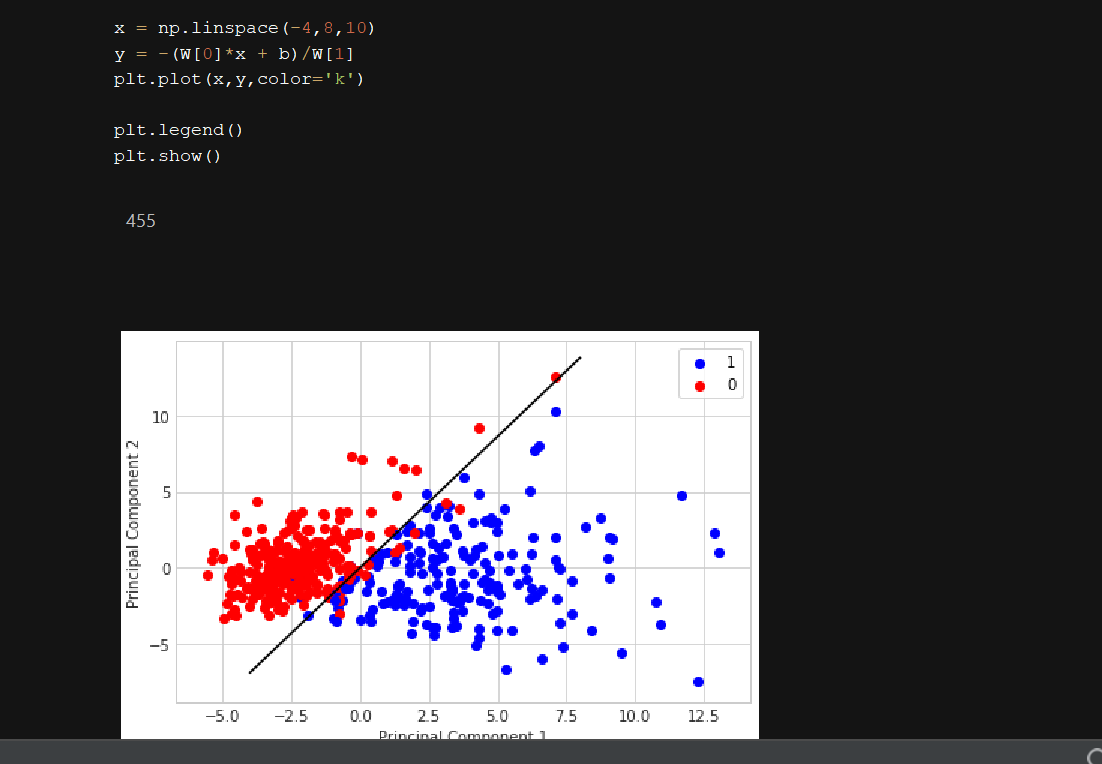
1. Print scatter plot of principal components



1. Perform PCR analysis and plot regression analysis







1. **Multinomial Logistic Regression**

Multinomial logistic regression is a [classification](https://en.wikipedia.org/wiki/Statistical_classification) method that generalizes [logistic regression](https://en.wikipedia.org/wiki/Logistic_regression) to [multiclass problems](https://en.wikipedia.org/wiki/Multiclass_classification), i.e. with more than two possible discrete outcomes.[[1]](https://en.wikipedia.org/wiki/Multinomial_logistic_regression#cite_note-1) That is, it is a model that is used to predict the probabilities of the different possible outcomes of a [categorically distributed](https://en.wikipedia.org/wiki/Categorical_distribution) [dependent variable](https://en.wikipedia.org/wiki/Dependent_variable), given a set of [independent variables](https://en.wikipedia.org/wiki/Independent_variable) (which may be real-valued, binary-valued, categorical-valued, etc.).

Multinomial logistic regression is used when the [dependent variable](https://en.wikipedia.org/wiki/Dependent_variable) in question is [nominal](https://en.wikipedia.org/wiki/Level_of_measurement#Nominal_measurement) (equivalently categorical, meaning that it falls into any one of a set of categories that cannot be ordered in any meaningful way) and for which there are more than two categories. Some examples would be:

* Which major will a college student choose, given their grades, stated likes and dislikes, etc.?
* Which blood type does a person have, given the results of various diagnostic tests?

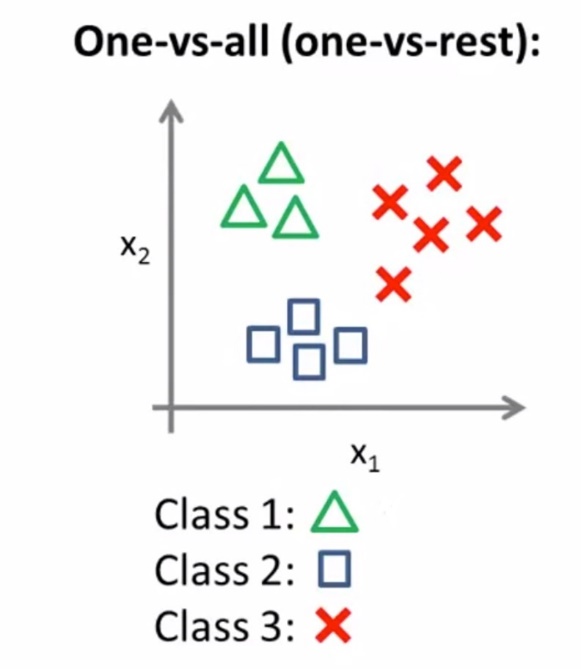
In the binary classification, logistic regression determines the probability of an object to belong to one class among the two classes.

If the predicted probability is greater than 0.5 then it belongs to a class that is represented by 1 else it belongs to the class represented by 0. In multinomial logistic regression, we use the concept of one vs rest classification using binary classification technique of logistic regression.

Now, for example, let us have “K” classes. First, we divide the classes into two parts, “1 “represents the 1st class and “0” represents the rest of the classes, then we apply binary classification in this 2 class and determine the probability of the object to belong in 1st class vs rest of the classes.

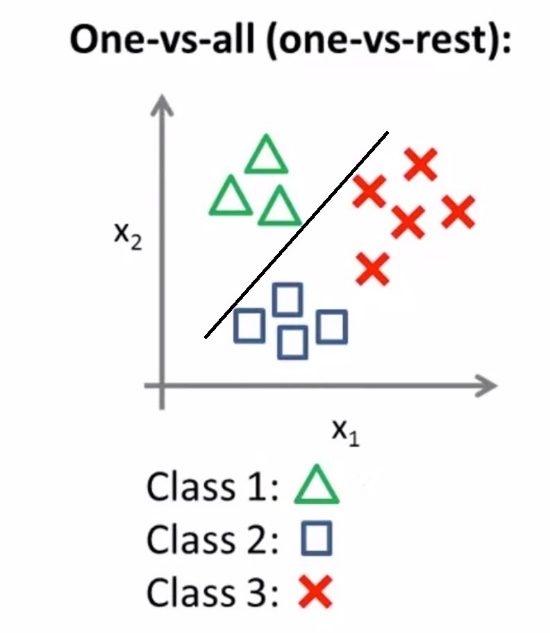
Similarly, we apply this technique for the “k” number of classes and return the class with the highest probability. By,  this way we determine in which class the object belongs. In this way multinomial logistic regression works. Below there are some diagrammatic representation of one vs rest classification:-

Step 1:-



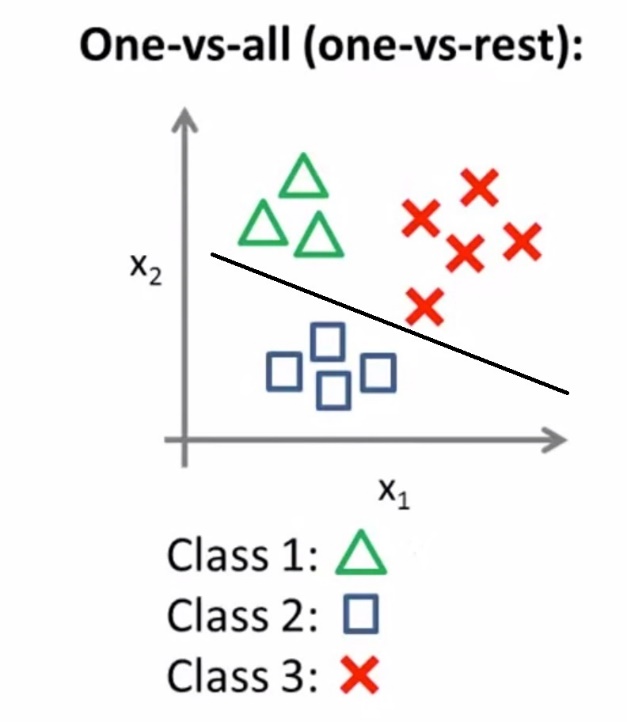
Here there are 3 classes represented by triangles, circles, and squares.

Step 2:



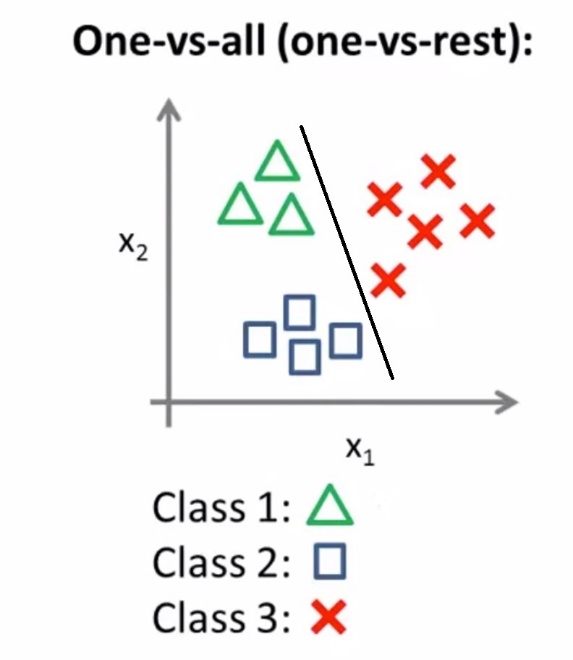
Here we use the one vs rest classification for class 1 and separates class 1 from the rest of the classes.

Step 3:

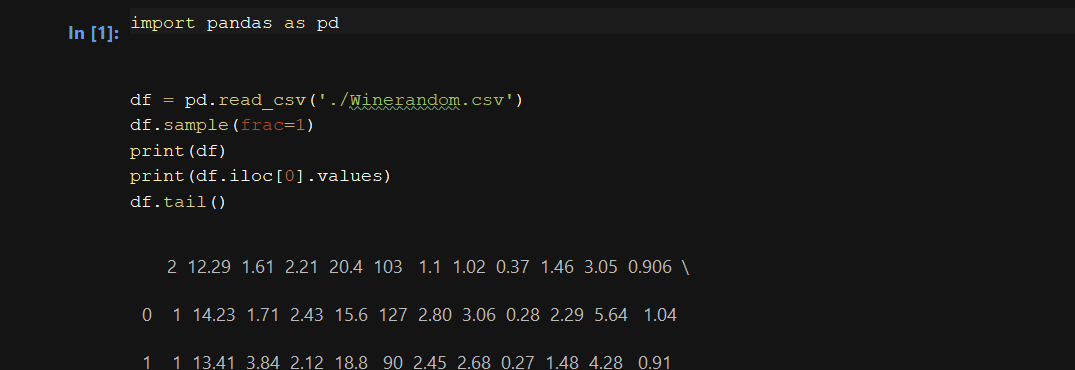


Here we use the one vs rest classification for class 2 and separates class 2 from the rest of the classes.

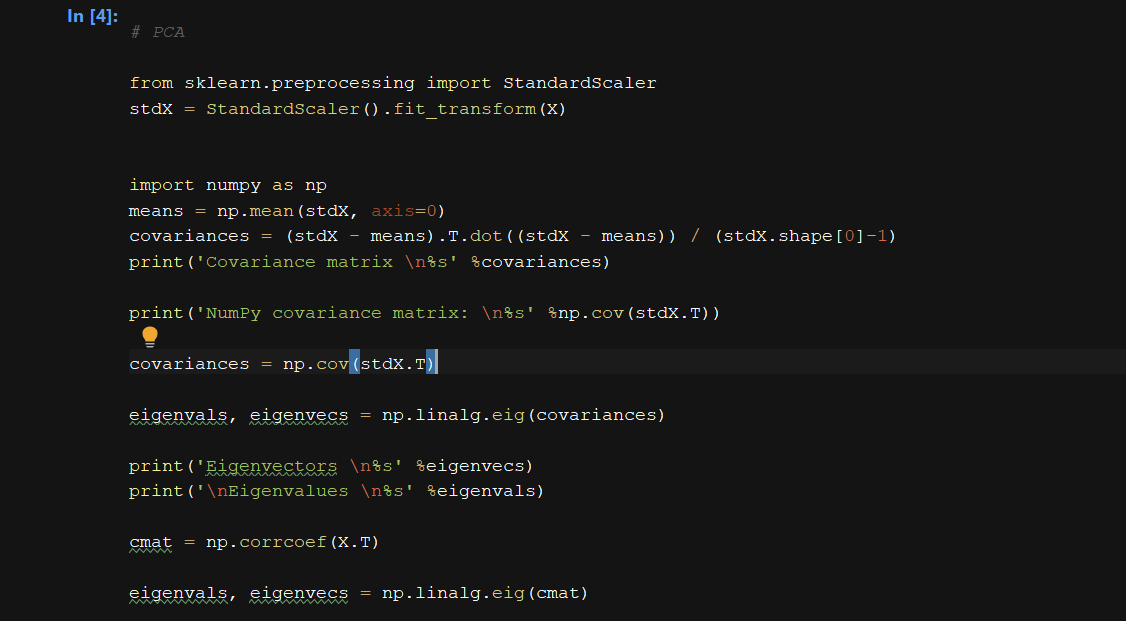
Step 4:



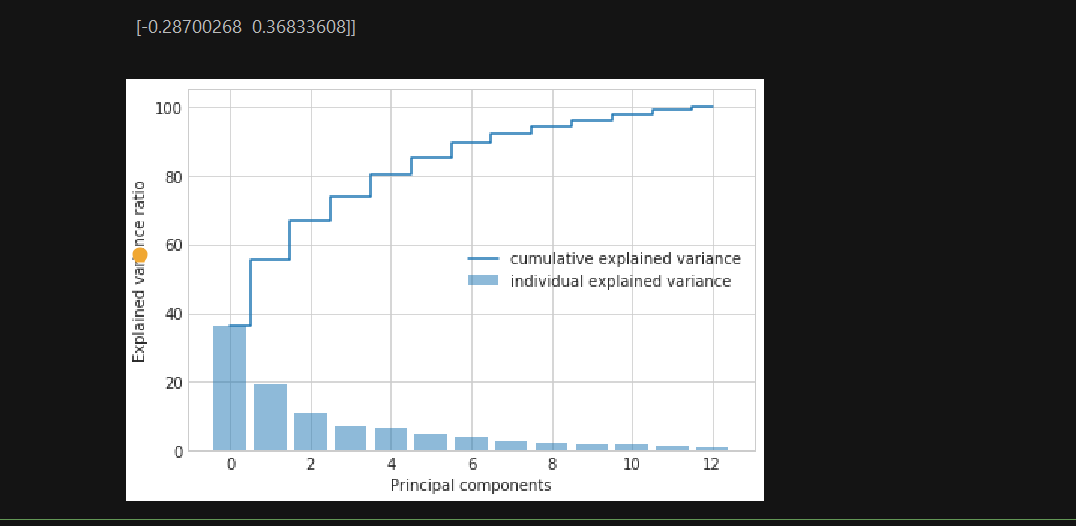
1. Dataset Used – Wines



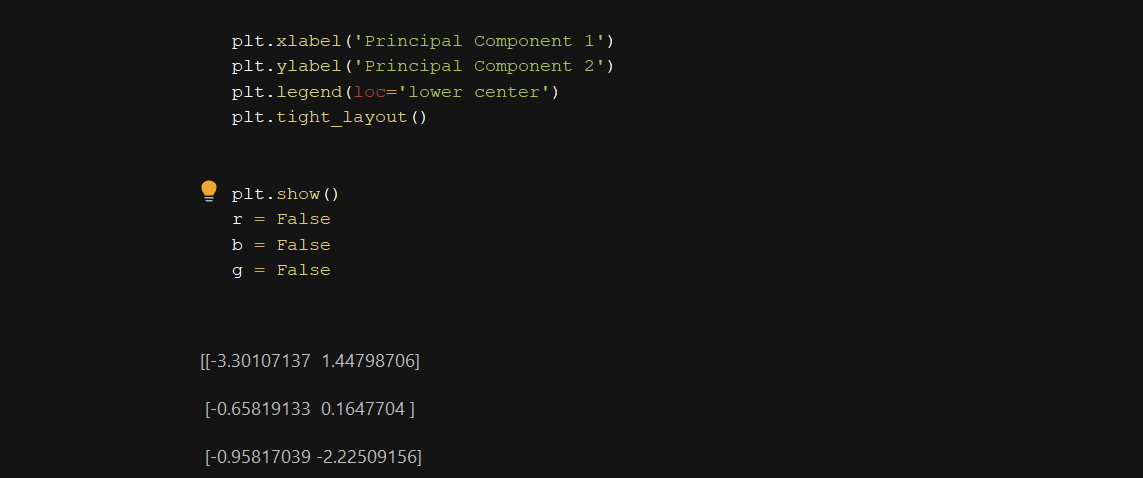
1. Perform PCA

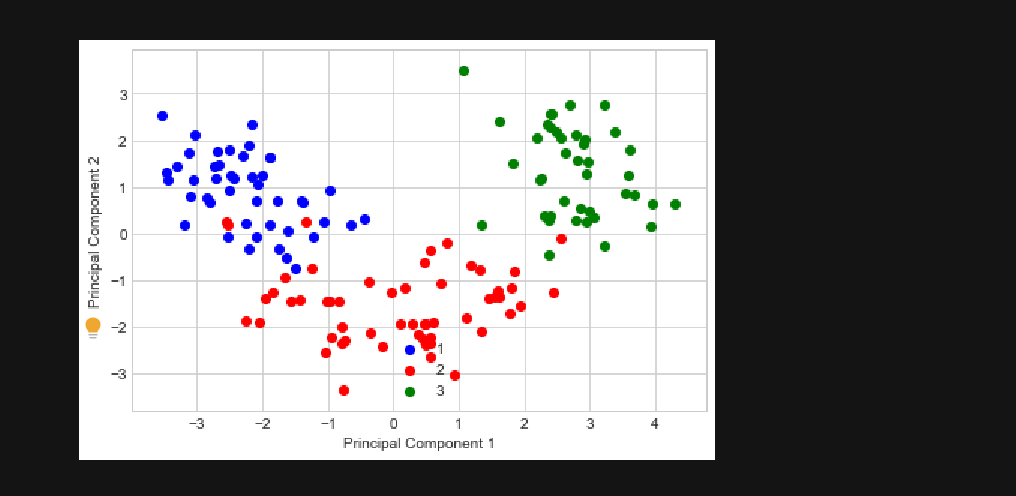


1. Plot co-variance ratio



1. Create scatter plot with principal components

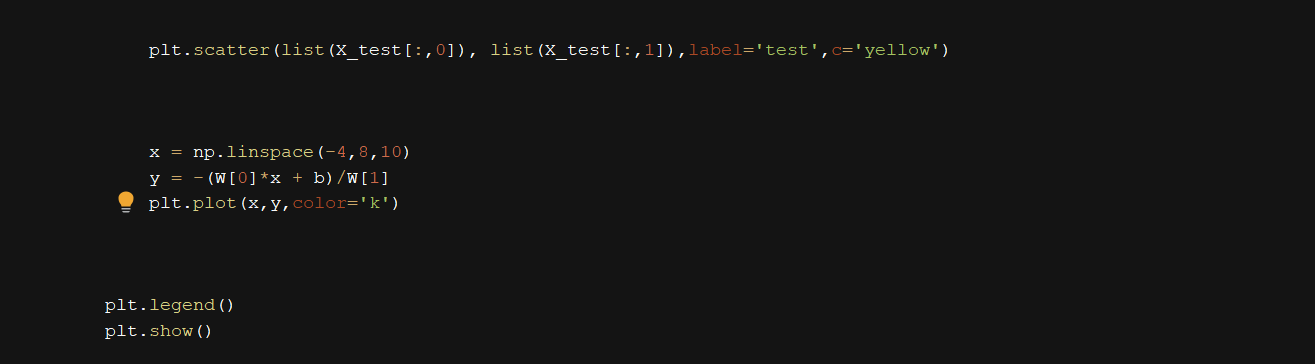


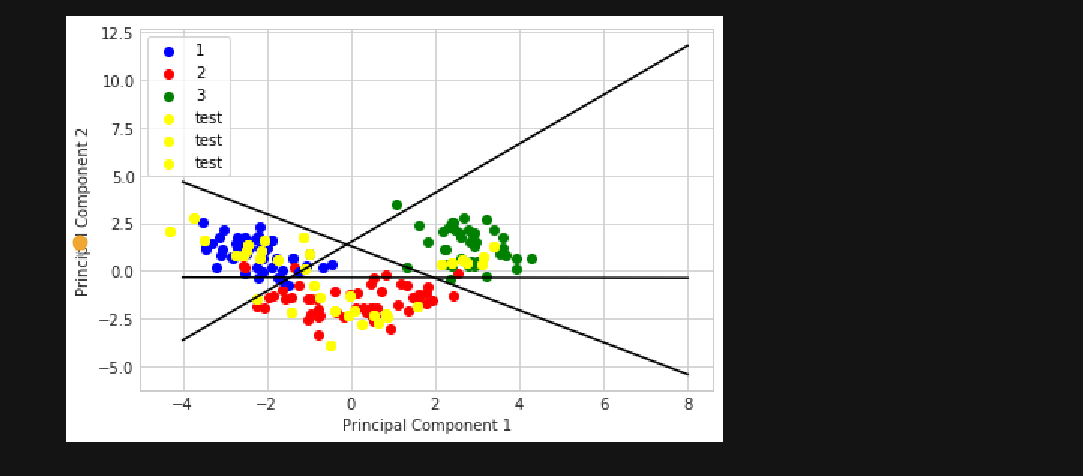


1. Perform multinomial regression analysis

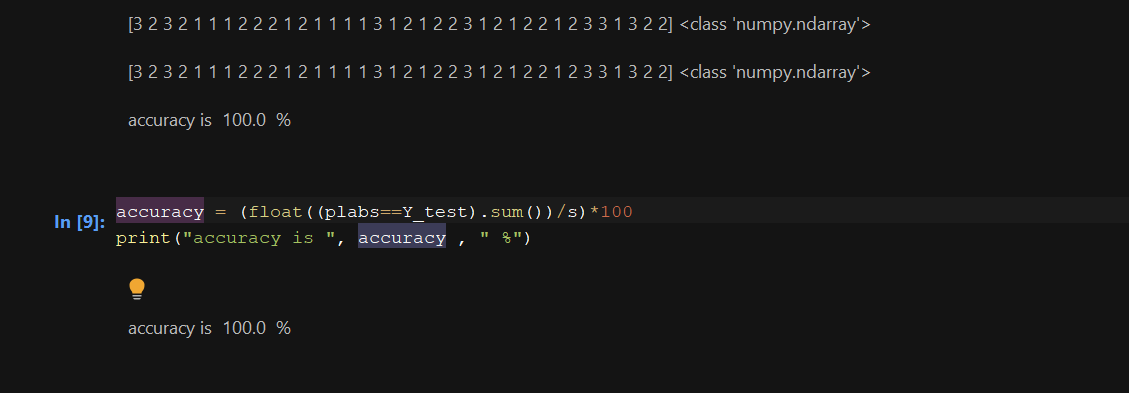


1. Create scatter plot with regression results

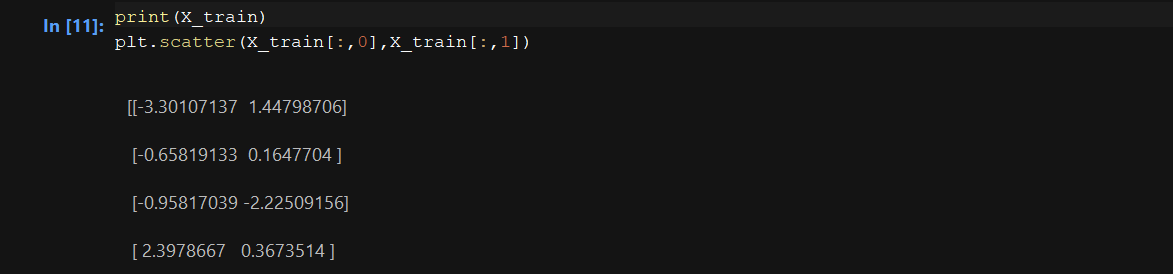


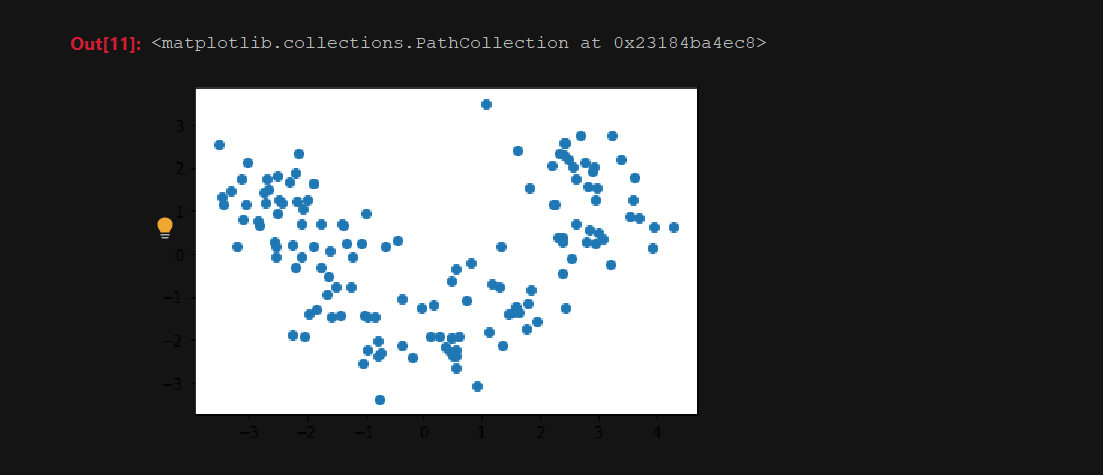


1. Results accuracy – 100%



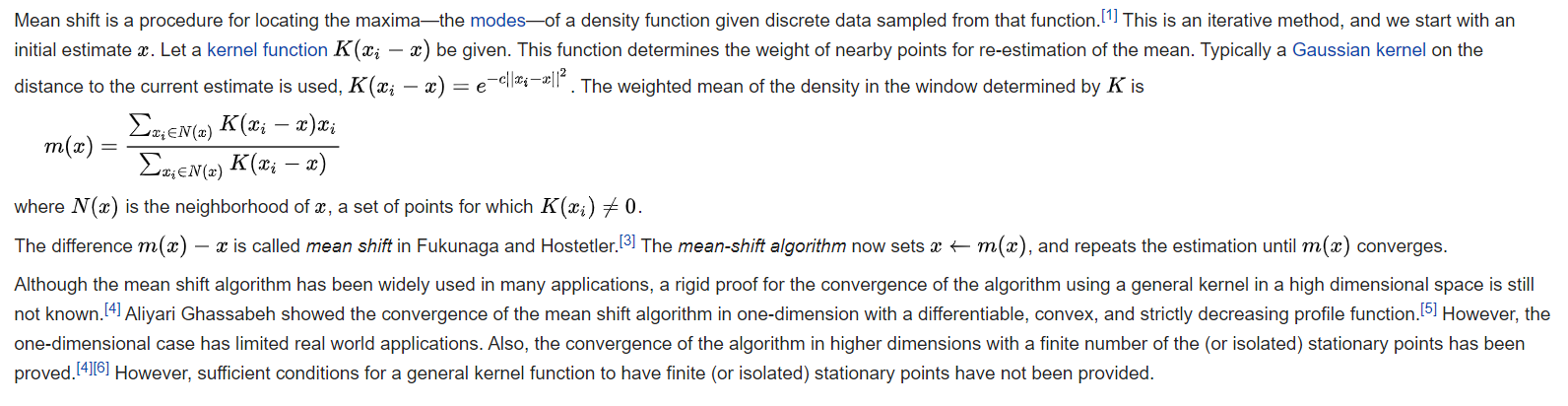






1. **Mean-shift clustering**

**Meanshift**is falling under the category of a clustering algorithm in contrast of Unsupervised learning that assigns the data points to the clusters iteratively by shifting points towards the mode (mode is the highest density of data points in the region, in the context of the Meanshift). As such, it is also known as the **Mode-seeking algorithm**. Mean-shift algorithm has applications in the field of image processing and computer vision. Unlike the popular K-Means cluster algorithm, mean-shift does not require specifying the number of clusters in advance. The number of clusters is determined by the algorithm with respect to the data

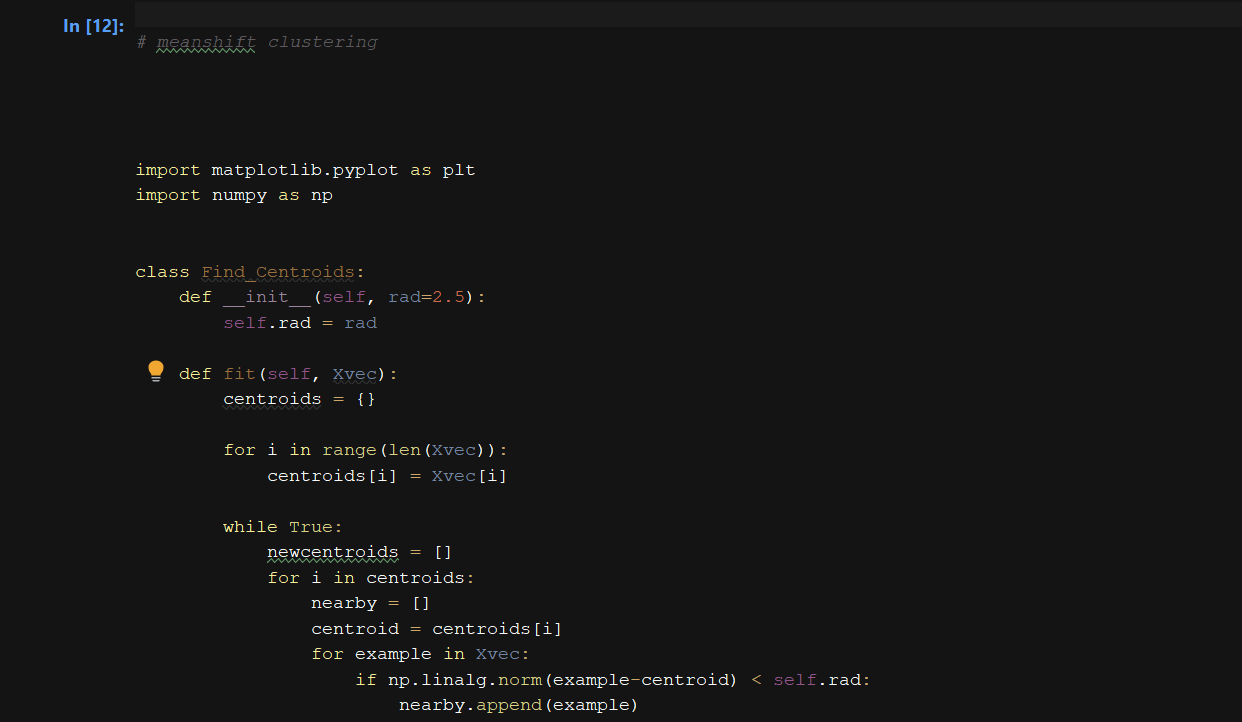


X-Means

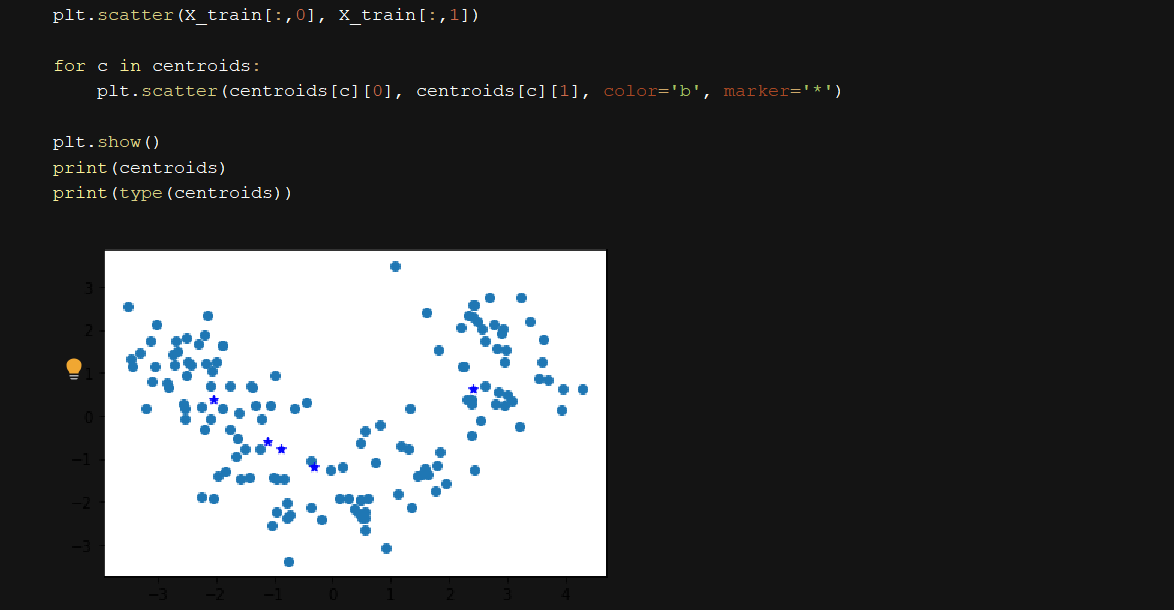
The obvious shortcomings of the basic k-means clustering are that the number of clusters needs to determined in advance and the computational cost with respect to the number of observations, clusters, and iterations.

What X-means does exactly is the following. Assuming a user supplied lower and upper bound for the number of clusters, it ﬁnds all locally optimal k-means clusterings within this range, evaluates them using the BIC criterion, and at the end returns the clustering that evaluates the best. X-means can be performed by splitting, or other random criteria. We have used mean-shift to find optimal number of cluseters

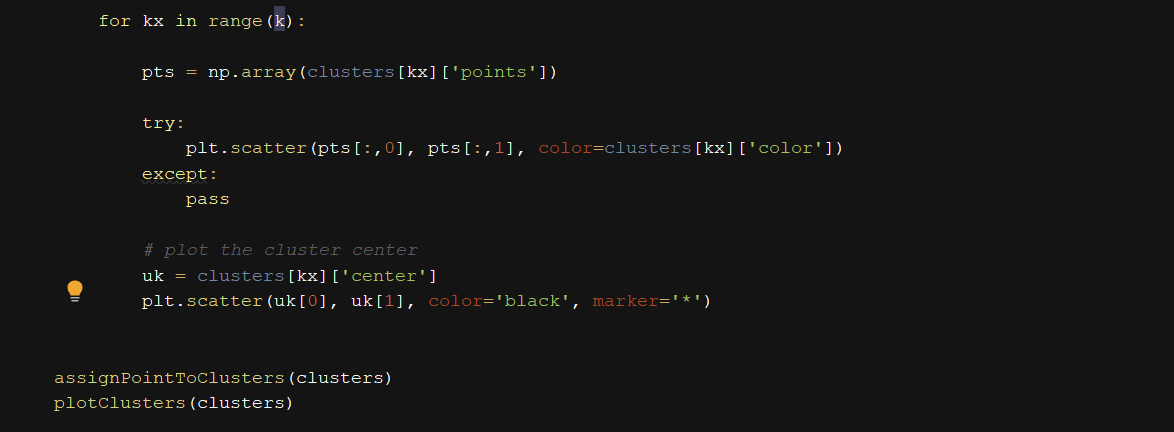
1. Finding centroids

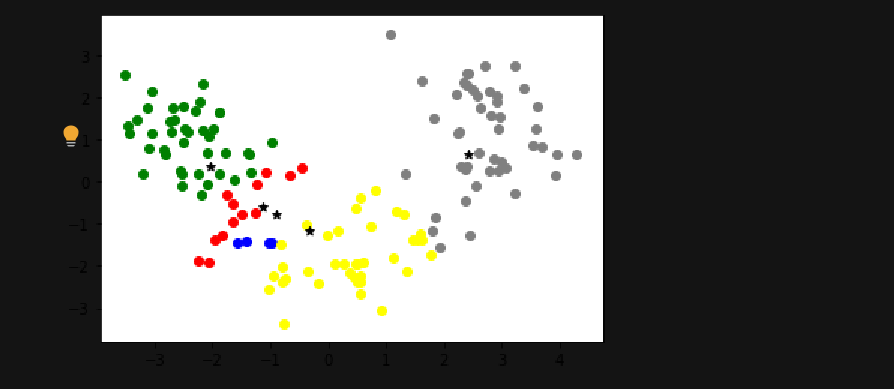


1. Print centroids



1. Assign points to clusters





1. Update Clusters

