

Karunachandra R.H.I.O.

E/17/153

CO 544 – Lab 07

a) Task 1

State in your report that you have completed all the four labs and one takeaway point for each lab in your own words.

Lab 01

Sampling and Projection

Sampling is a process we can use in statistical analysis. When there is a population of data, we extract some predetermined dataset from that population. This is called sampling. Projection is kind of cartesian visualization. When the dimension gets higher it is difficult to observe data. Projection will reduce the dimensions and provide an easy visualization of data. This is a mainly used technique in PCA (Principal Component Analysis) for dimensional reduction.

Lab 02

Perceptron

Classification problem of two classes was done, as how all the one class is having positive dot product and the other class's points are having negative dot product with a particular direction. Perceptron algorithm is used here.

We randomly sampled data from the given dataset and checked whether it is correctly classified and if it is we do nothing. If it is misclassified, we update. As we are doing that, we monitored the error on the training set and an unseen test set to check how well the model is performing on that particular data set.

After finding data points are linearly separable, we make them closer until some percentage of them are overlapping on each other. Then we check out what happens when data points are on (1,1) line.

Lab 03

Regularization

When we have very large number of data when compared to the dimensions, things are straight forward. But if we have a small amount of data and the dimensions are very high, we have to use regularization. We control how complex the model can be by regularizing. Basic idea of this is coming from linear regression.

In this lab, we looked into two main regularizations: quadratic and sparsity induced.

Quadratic regularization's intention is to produce smooth models. So, we don't allow the parameters to go to very large values. They are controlled as what values they can be. This scenario is achieved by quadratic regularization.

Sparsity inducing regularization look at all the weights that are coming out and push small ones towards zero. So, we end up with a model which has a small number of parameters.

Lab 04

Histogram

Data visualization is the main thing discussed in this lab.

Histogram is used in frequency distributions of set of continuous data for plotting purposes. As the convention it's better to create the plot first. So that, we can mention the height (second parameter) and width (first parameter) of the figure. When we plot, it's good to always go with a perfect aspect ratio. Aspect ratio means basically the ratio between the width and the height of the figure.

```
plt.figure(figsize=(10, 6))    #Create a figure  
plt.hist(recent_year_df['gdpPercap']);    #Plot the histogram
```

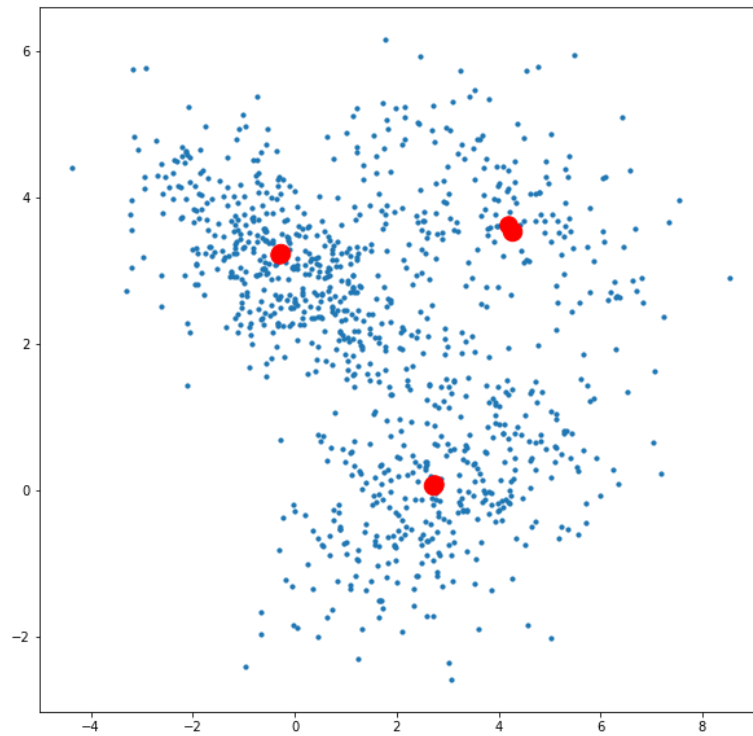
These are the commands for plotting the histogram. In the second command if we miss the semicolon, it will also give the resulting plot with some printed values. To avoid those values, we have to put a semicolon at the end.

In histograms, each column represents a bin. When we draw a histogram by default bin value is 10. But we can change those things as our preference.

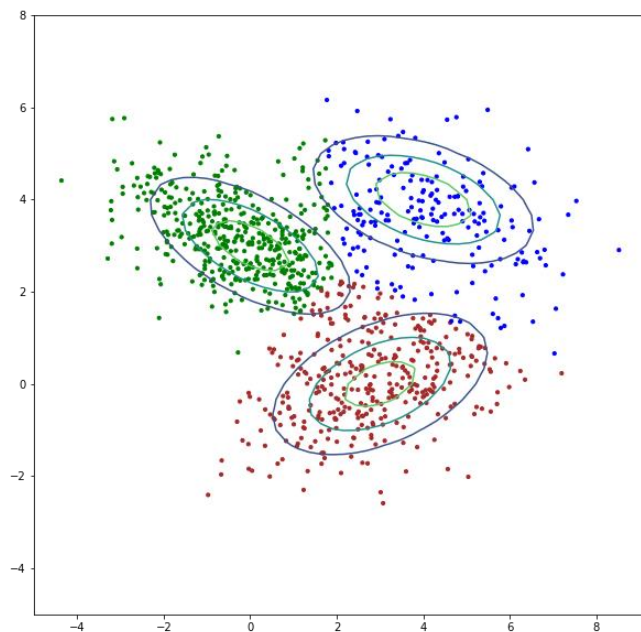
b) Task 2

https://colab.research.google.com/drive/1_mAUBNtYd-Gai0eHVBxwEVurVmaEWSnh#scrollTo=cxRLNDHknXIM

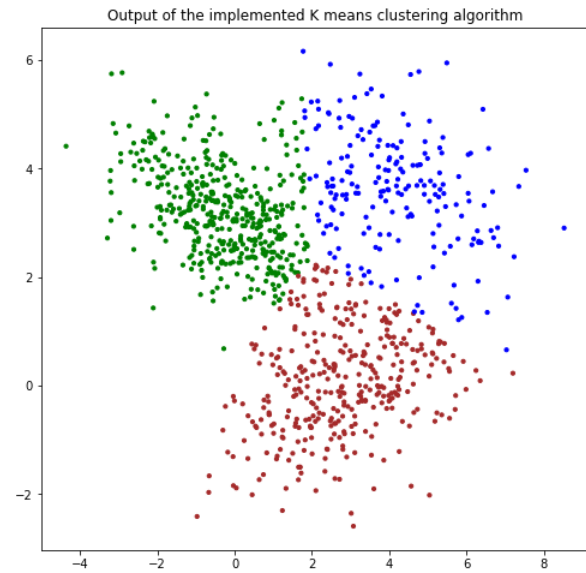
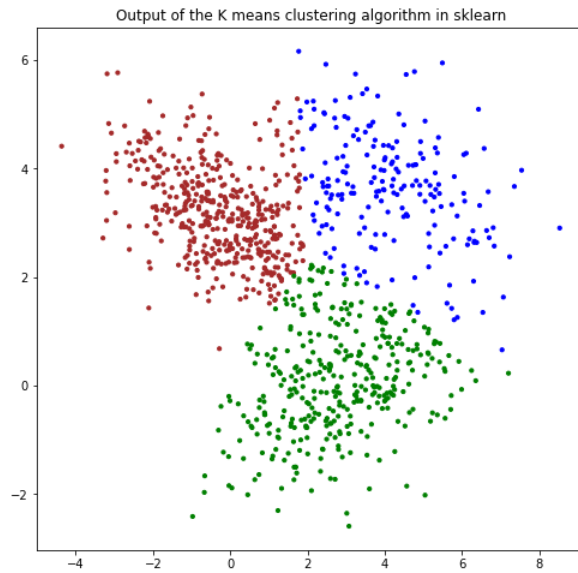
1.



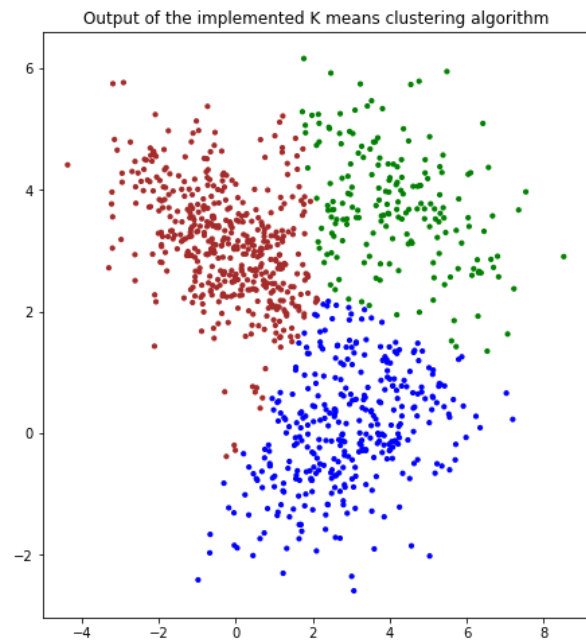
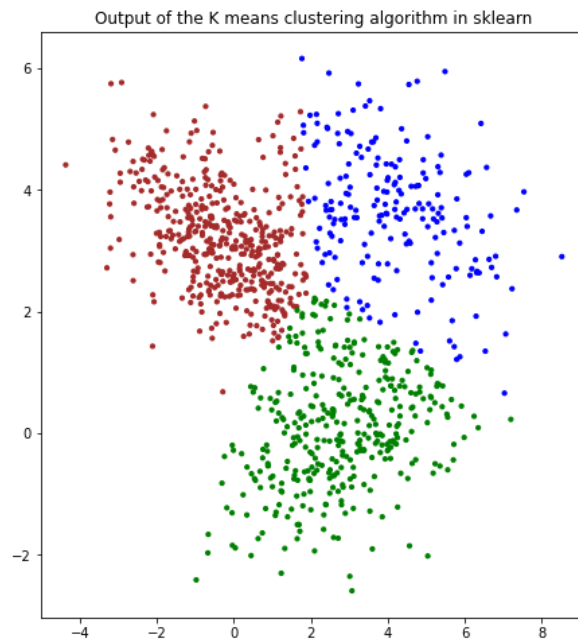
2.



3.



4. Algorithm failing



5. Here selected the iris.csv dataset

