Free Sound General-Purpose Audio Tagging

Bohan Li
Department of EECS
University of Tennessee, Knoxville

Pengxiang Xu
Department of EECS
University of Tennessee, Knoxville

Abstract

Automated universal audio tagging system created in this project can recognize some sounds in our daily life by a sort of methods including PCA, Gaussian, K-means, K-NN and SVM. Training and testing is based on the the dataset provided in the Free Sound Universal Audio Marking Challenge. After training more than 500 manually annotated audio events, results show that SVM has the best performance of classification with 75 % while K-means has the lowest accuracy of 3%.

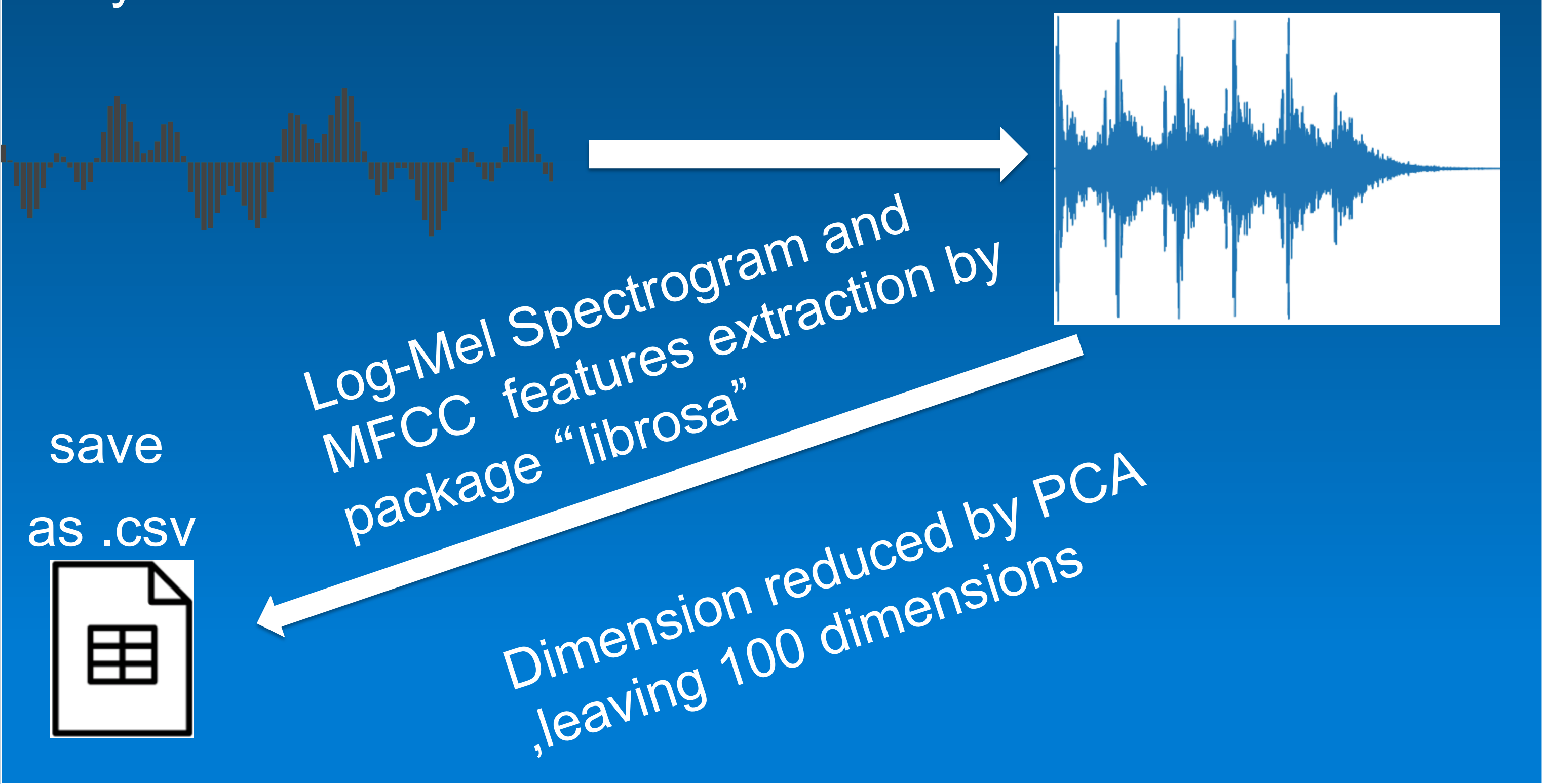
Introduction

As human, we cannot recognize the daily sounds because it's kind of wave and usually transimitted into ear. Currently, sounds collection and audio precossing cost a lot so that sound recognition based on speech technology become a high-tech technology. To do a better recognition, data should be deal with following steps:

- First, the data set is audio and needs to convert the audio into processable data.
- Second, the audio of the data set contains noise and needs to solve the error caused by noise.
- Third, the audio length It is not fixed, it need to be converted different lengths of audio to the same feature size.

Methodology

Daily sounds' wave records

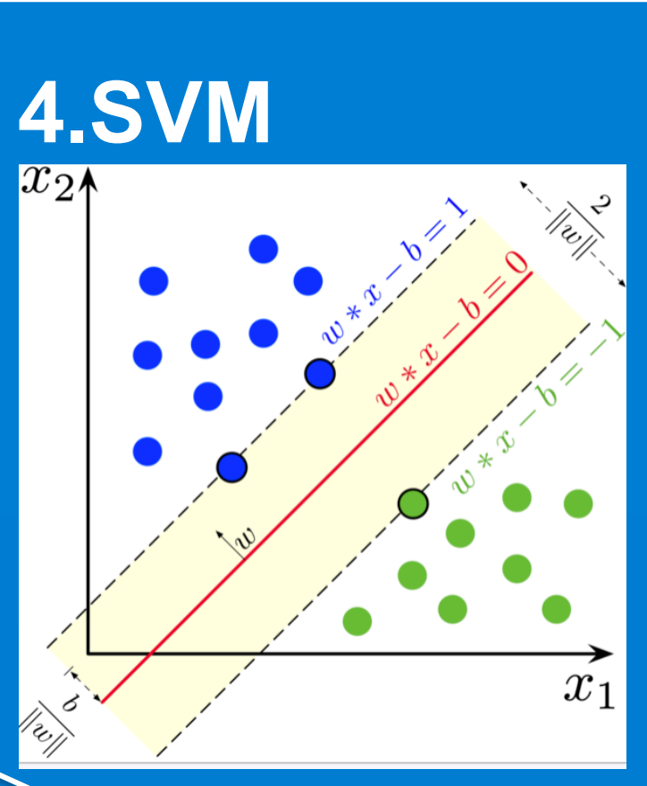


Log-Mel Spectrogram and MFCC features extraction by package "librosa"

save as .csv

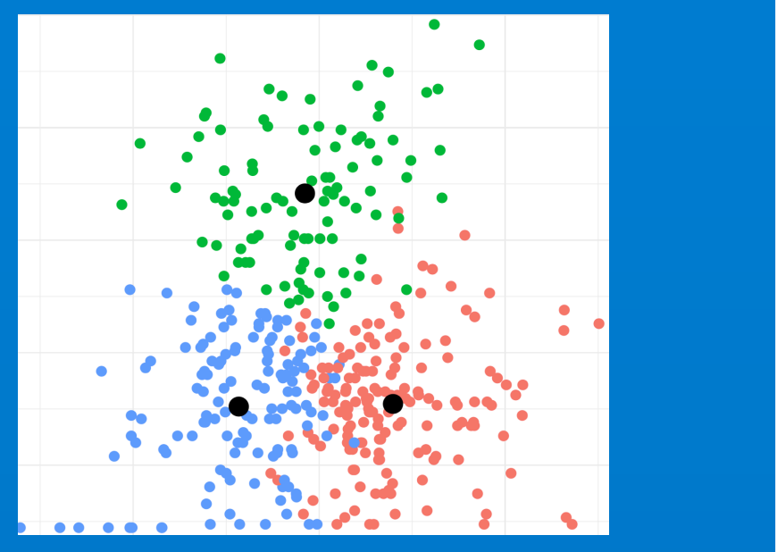
Dimension reduced by PCA ,leaving 100 dimensions

4.SVM



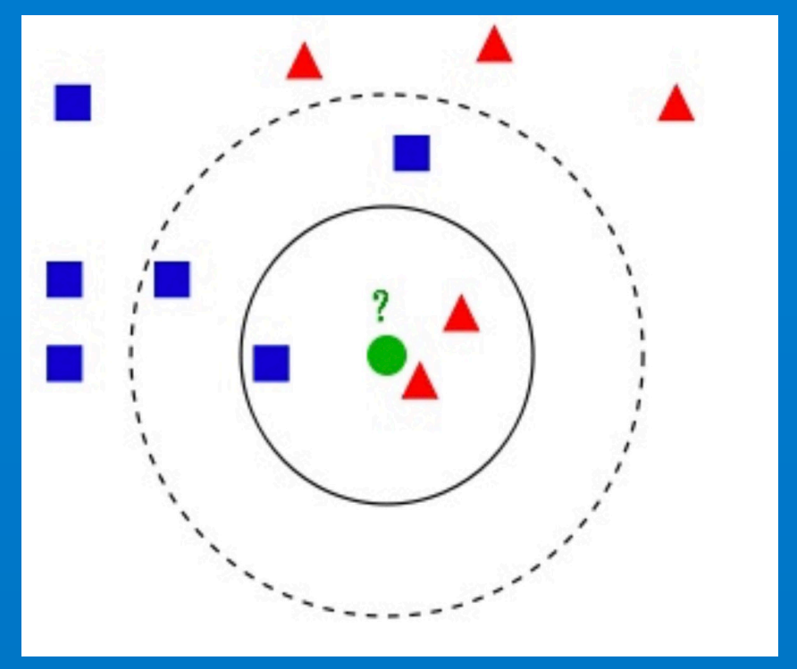
- One-versus-rest
- Test penalty c from: [0.0001, 0.001, 0.01, 0.1, 1], then use the best value to do the classification
- m-fold: 3-10

3.KNN



- Reduce to 2 dimensions
- Classify into 41 clusters

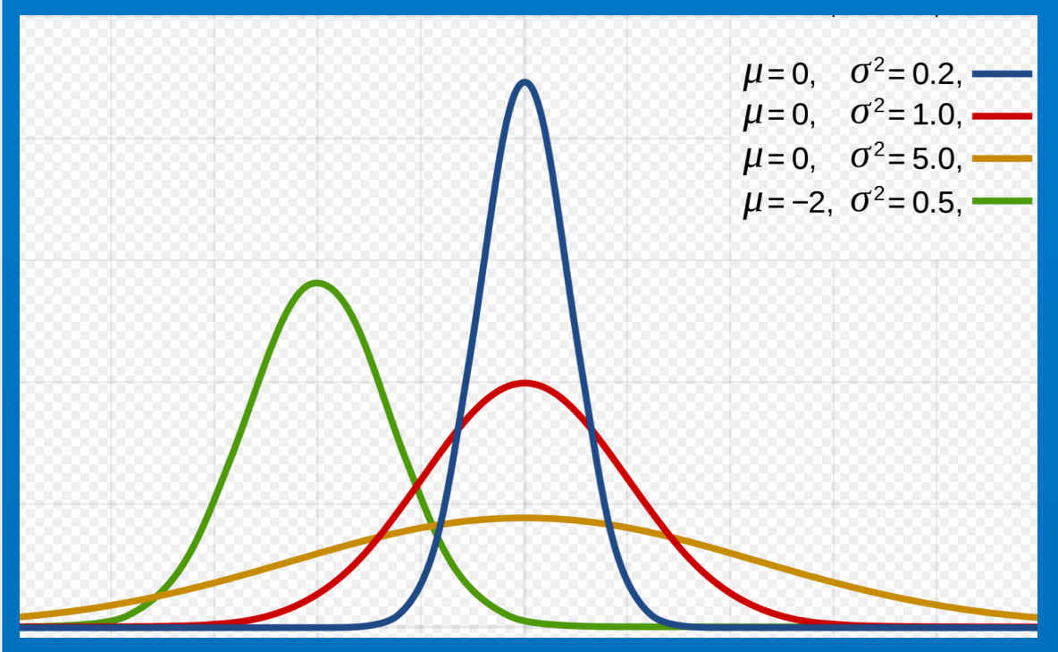
2.K-mean



- Value of K: 1-10
- Euclidean Distance for each dimension

$$E(x,y) = \sqrt{\sum_{i}^{n}(x_i - y_i)_2}$$

1.Gaussian



- Transfer 41 labels into 41 numbers
- Calculate mean and covariance for each class
- We have 41 classes, so we build 41 Gaussian formula for each case
- Put into test data and get the predict label

Results

- 1.PCA

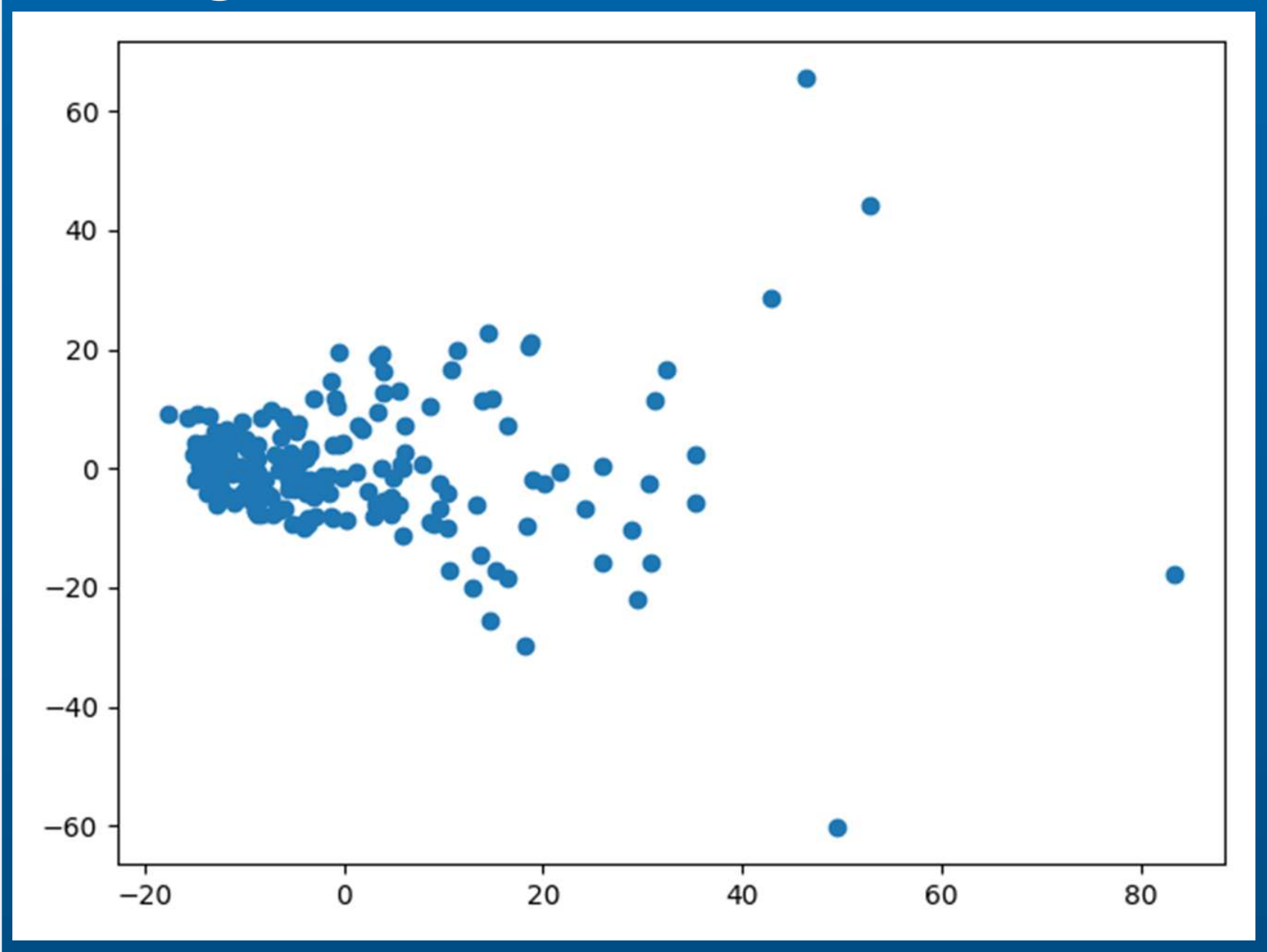


Figure 1. The First two dimensions of PCA

2.Gaussian

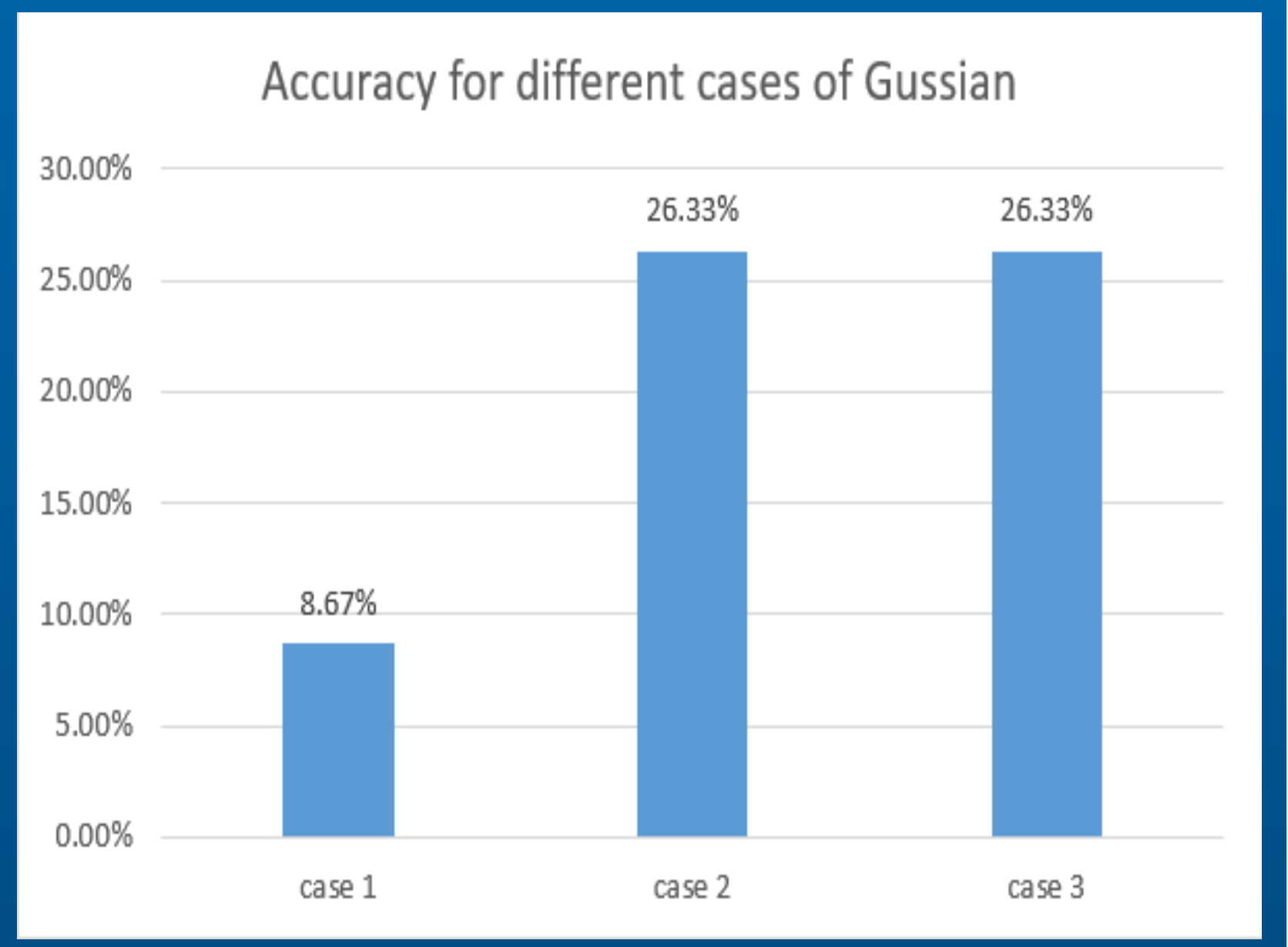


Figure 2. Three cases of Gaussian

3.K-NN

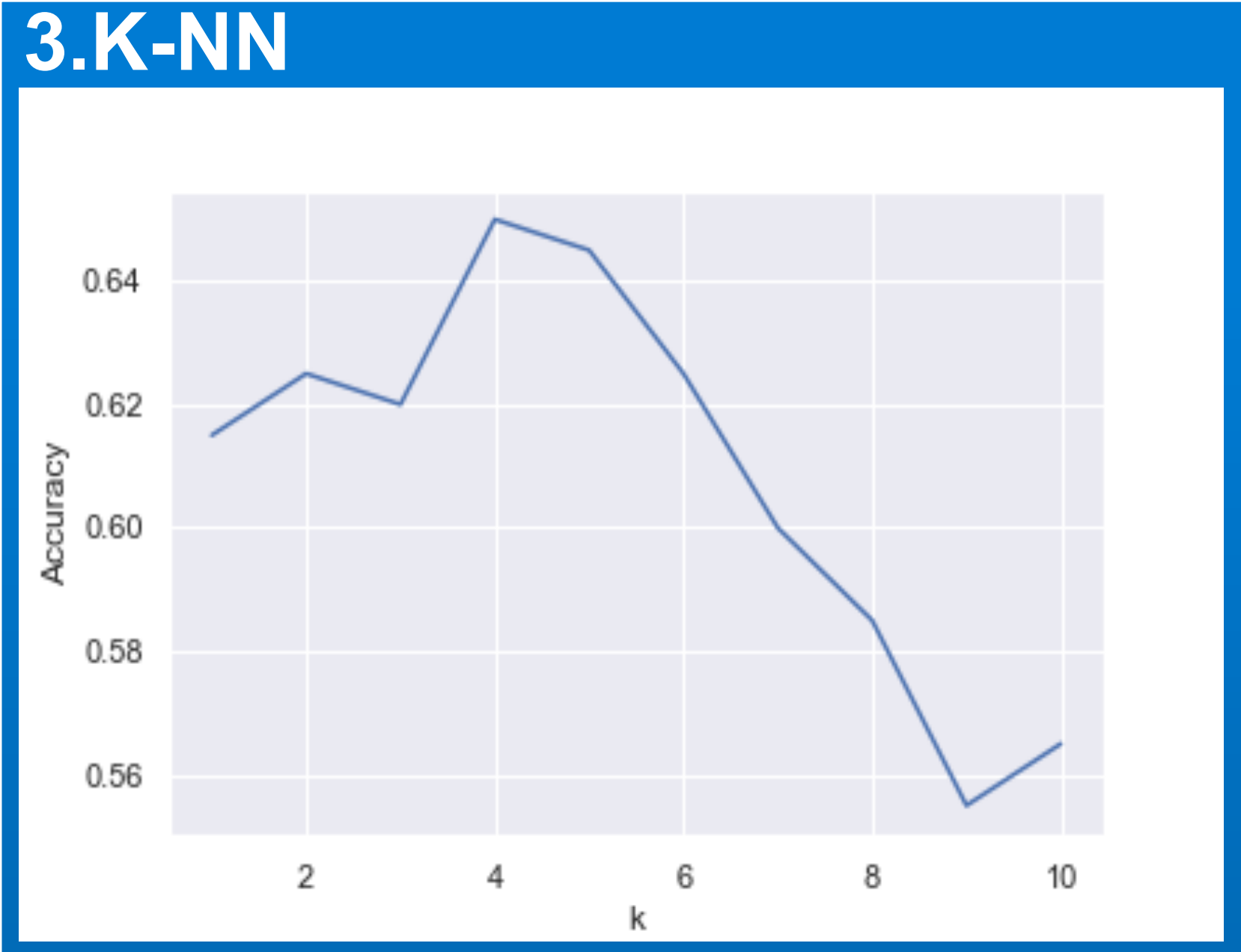


Figure 4. When the k is 4, KNN has the highest accuracy over 65%.

4.K-means

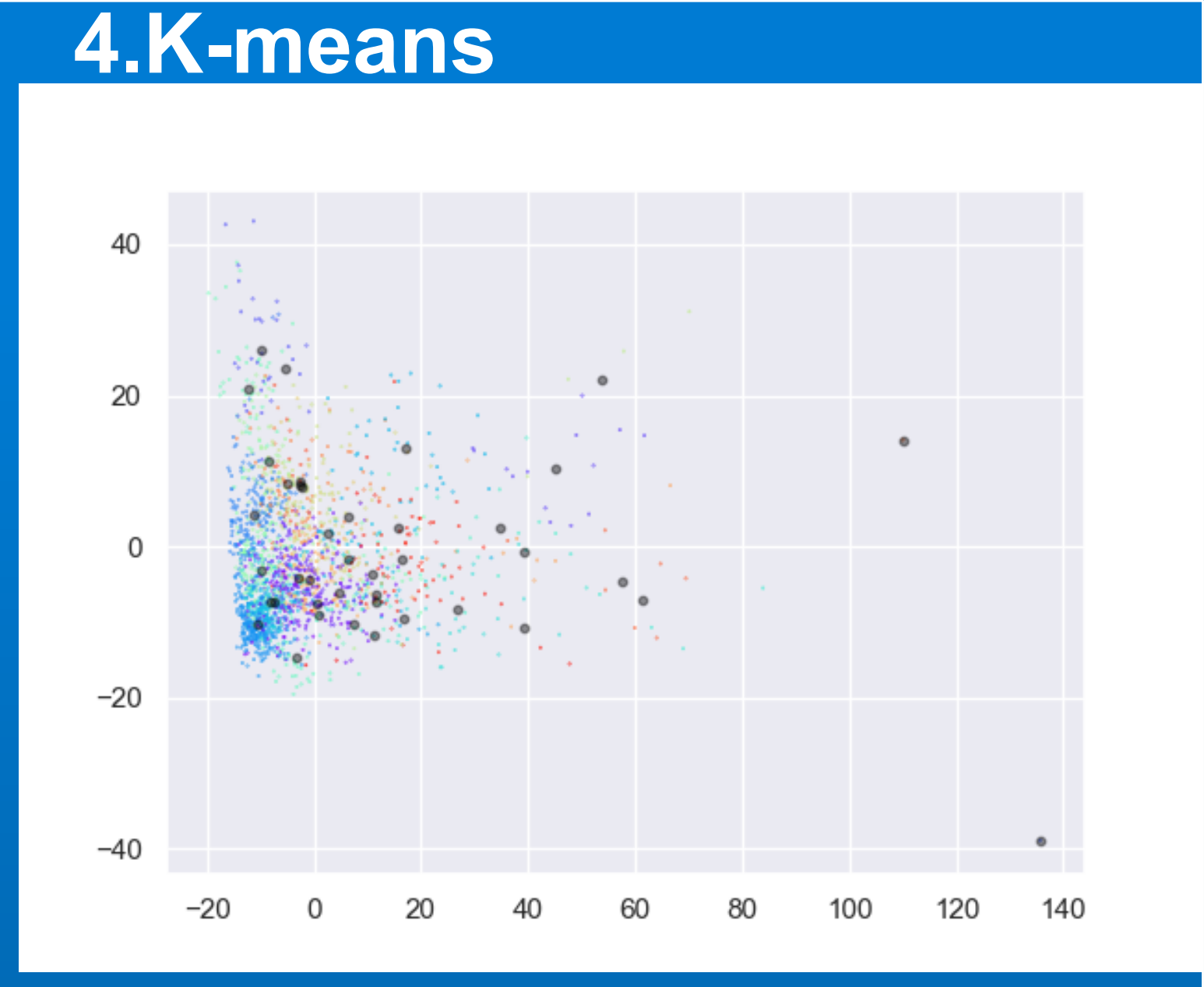


Figure 5. K-means has the worst accuracy that is less than 3%.

5.SVM

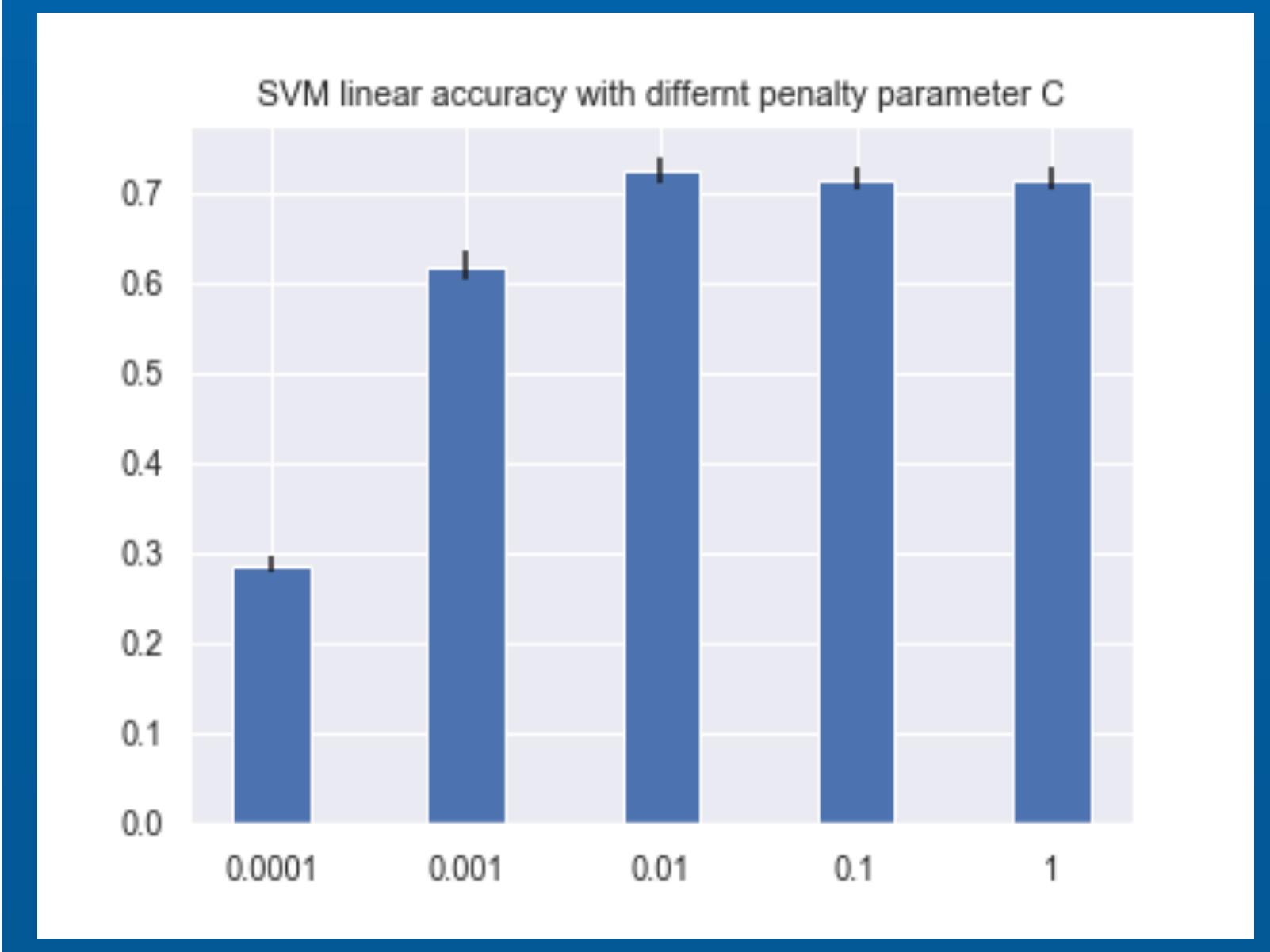


Figure 7. SVM has the highest accuracy almost 75%.

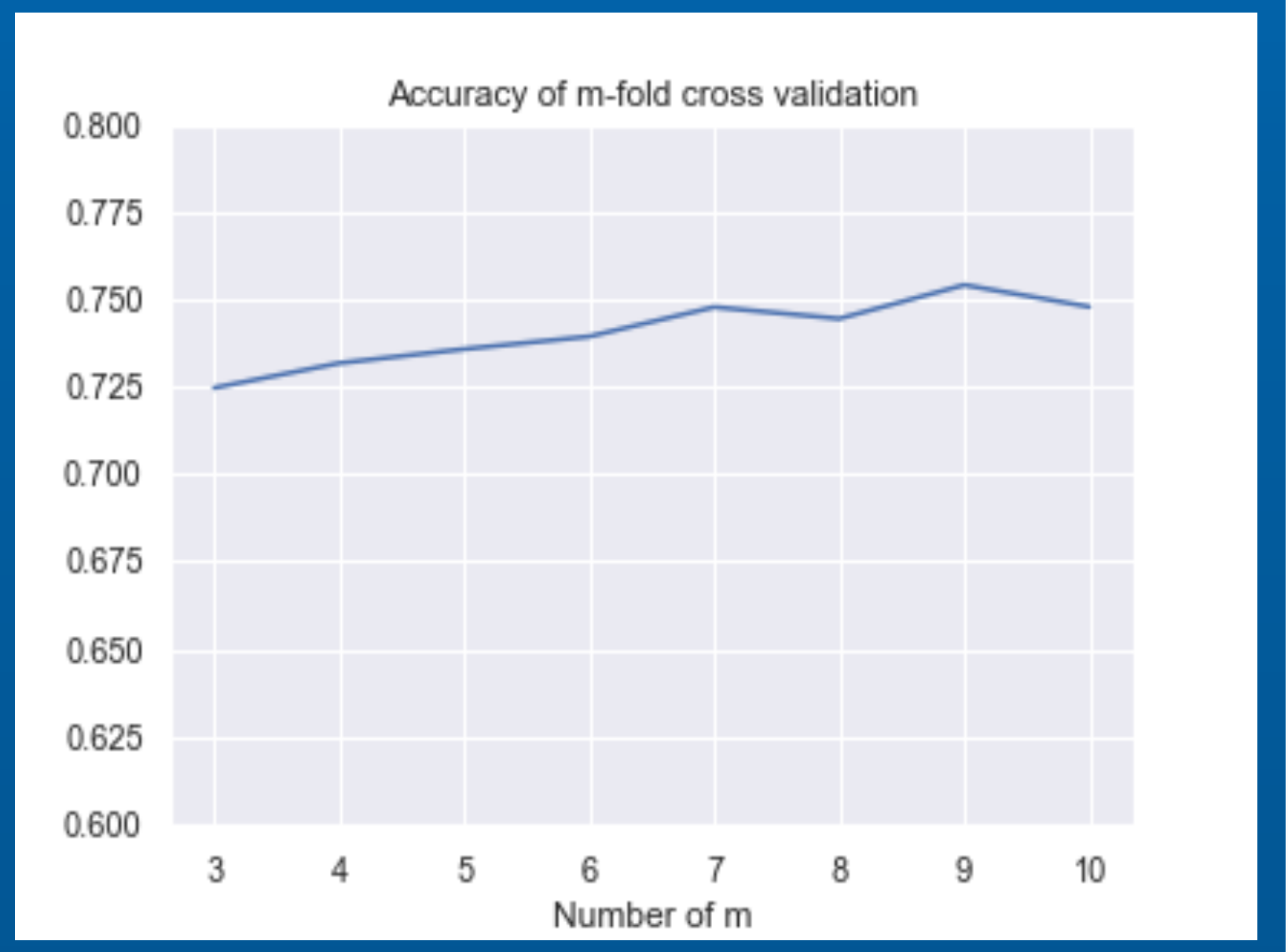


Figure 8. The best value of penalty C is 0.01.

Conclusion

- SVM has the highest accuracy while k-means has the lowest. That's mainly because even dimension is reduced, it's still large and many information could be lost during this process.
- Moreover, the accurarcy never reach 80% no matter how many diemensions are.
- Finally, not a desent dataset may casue this problem as only 3000 samples are used in this project

Acknowledgements

We would like to thank:

- Our mentor Dr.Toufer for her support and instruction of this final project.
- All teaching assistants for their instruction and help
- All classmates in COS526 for their discussion and communication

Reference

[1] https://en.wikipedia.org/wiki/Curseofdimensionality
[2] Rasmussen, Carl Edward. "Gaussian processes in machine learning." Summer School on Machine Learning. Springer, Berlin, Heidelberg, 2003.
[3] https://en.wikipedia.org/wiki/Support-vector_machine.
[4] S.-Y. Chou, J.-S. R. Jang, and Y.-H. Yang. Learning to recognize transient sound events using attentional supervision. In IJCAI, pages 3336–3342, 2018.
[5] Vishwanathan, S. V. M., and M. Narasimha Murty. "SSVM: a simple SVM algorithm." Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290). Vol. 3. IEEE, 2002.
[6] Michie, Donald, David J. Spiegelhalter, and C. C. Taylor. "Machine learning." Neural and Statistical Classification 13 (1994).
[7] Bishop, Christopher M. Pattern recognition and machine learning. springer, 2006.