

Gaussian Mixtures in Machine Learning

Miguel Antonio Rodriguez Delgado
Dortmund, Germany
miguel-antonio.rodriguez-delgado@stud.hshl.de

Abstract—Gaussian or Normal models help us to represent most of the distribution of all aspects of the world, showing how they can spread regarding the possibility for it to happen. This Gaussian models can allow us to distinguish between different set of points in an image for example. But in certain cases a point can be part of different groups. In this case Gaussian Mixtures can allow us to represent patterns which have elements in common. Gaussian Mixtures can be used in different fields. For example, to generate cancer detection algorithms.

Index Terms—Gaussian Mixtures, Deep learning, Machine learning.

I. INTRODUCTION

In recent years there have been an increasing interest in deep learning algorithms for clustering detection, And deep neural networks have improved in the field of supervised classification [3]. Some important examples of this are present in Face-book deep face software detection [3], prediction of pancreatic [2] and liver [5] cancer, among others.

Since Gaussian Mixtures are based on a combination of Gaussian Normal models, it is important to understand what is a Gaussian Normal mode (section II). Then Gaussian Mixtures will be discussed (Section III) and finally some examples on how Gaussian Mixtures can be applied on different research areas.

In [4] and [3] we find a description on how Gaussian mixtures work in a general way. In reference [1] is described how deep learning algorithm can be represented as Gaussian mixtures. And in [2] and [5] we can find examples on how to use Gaussian Mixtures to detect and predict pancreatic and liver cancer.

II. GAUSSIAN MODELS

Most of the parameters related on classification can be represented thought Gaussian Models [6]. An simple example of this is the height of the population. In Fig II we can see a representation of this. If we start analysing the figure from left to right we will see that the amount of short people is small, since is not very frequent to find small people. As we move on towards the middle of the graph, the amount of people start increasing, showing that most of the population will be present in this central part of the table. A proof of this in the real world, is that most of the people that we interact with is going to be near this height. But as we keep moving more to the right side of the graph, it will decrease again. Here once again we can see that the amount of tall people in our surroundings is not very high. This distribution will give us the shape of a bell. An this bell shape will be present in any distribution that we want to analyse.

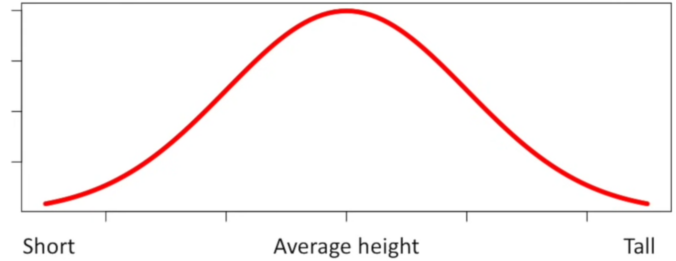


Fig. 1. Gaussian distribution of the population height. StatQuest with Josh Starmer. The Normal Distribution, Clearly Explained!!!. 2017.

Another example will be regarding the times on a race II. The best racers will be on the left of the graph, that will represent the one or two racers that will pass the goal at first. Then while the race time increases there is more people that will arrive together, in this case not only one or two at the same time but maybe hundreds, and at the end of the graph we will find those that are less prepared and again at the end of the racers, it will pass only a few, normally just one or two.

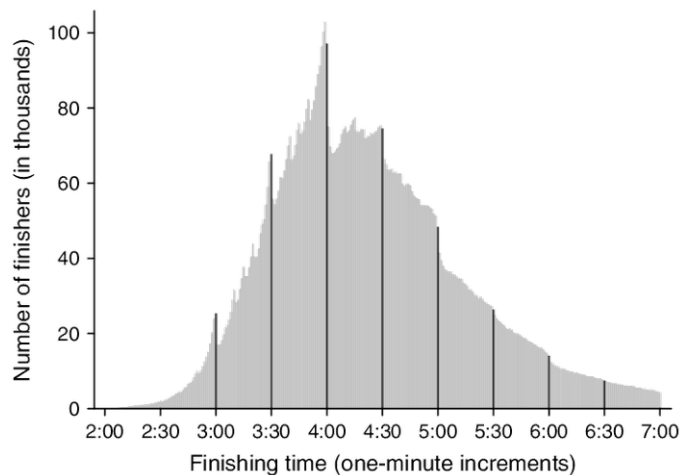


Fig. 2. Distribution of Marathon Finishing Times. Allen, Eric and Dechow, Patricia and Pope, Devin and Wu, George. (2016). Reference-Dependent Preferences: Evidence from Marathon Runners.

However this models will be capable to reproduce only single classification models.

III. GAUSSIAN MIXTURES

Since most of the models that have to be analysed contain more than one single distribution, a mixture of these distributions is required.

Gaussian Mixture Models are used for clustering applications. Some examples of this are audio classification, that would allow to determine different

REFERENCES

- [1] Seddik, M.E.A., Louart, C., Tamaazousti, M., & Couillet, R.. (2020). Random Matrix Theory Proves that Deep Learning Representations of GAN-data Behave as Gaussian Mixtures. *in Proceedings of the 37th International Conference on Machine Learning*, in *Proceedings of Machine Learning Research*, 119:8573-8582 Available from <https://proceedings.mlr.press/v119/seddik20a.html>.
- [2] Sekaran, K., Chandana, P., Krishna, N.M. et al. Deep learning convolutional neural network (CNN) With Gaussian mixture model for predicting pancreatic cancer. *Multimed Tools Appl* 79, 10233–10247 (2020). <https://doi.org/10.1007/s11042-019-7419-5>
- [3] Viroli, C., McLachlan, G.J. Deep Gaussian mixture models. *Stat Comput* 29, 43–51 (2019). <https://doi.org/10.1007/s11222-017-9793-z>
- [4] Nalisnick, E.T., Hertel, L., & Smyth, P. (2016). Approximate Inference for Deep Latent Gaussian Mixtures.
- [5] Amita Das, U. Rajendra Acharya, Soumya S. Panda, Sukanta Sabut, Deep learning based liver cancer detection using watershed transform and Gaussian mixture model techniques, *Cognitive Systems Research*, Volume 54, 2019, Pages 165-175, ISSN 1389-0417, <https://doi.org/10.1016/j.cogsys.2018.12.009>. (<https://www.sciencedirect.com/science/article/pii/S1389041718310143>)
- [6] The Normal Distribution, Clearly Explained!!! (2017, 9 oktober). YouTube. Geraadpleegd op 29 april 2022, van <https://www.youtube.com/watch?v=rzFX5NWojp0>
- [7] Allen, Eric and Dechow, Patricia and Pope, Devin and Wu, George. (2016). Reference-Dependent Preferences: Evidence from Marathon Runners. *Management Science*. 63. 10.1287/mnsc.2015.2417.
- [8] Gaussian Mixture Models. (2020, 28 december). YouTube. Geraadpleegd op 29 april 2022, van <https://www.youtube.com/watch?v=q71Niz856KE>