Basic Information

Projection Pursuit density estimation
Author: Jerome H. Friedman, Werner Stuetzle (Stanford) & Anne Schroeder
Published: Journal of the American Statistical Association, Volume 79, 1984 Pages 599-608
Citation: 305

Summary

The author introduced a novel method from projection pursuit to solve the nonparametric multivariate density estimation problem, mathematically, to estimate the probability density of a p-dimensional random vector $X \in R^p$ on the basis of iid observations, without assumptions of parametric data. The previous methods all failed to address this problem because of the curse of dimensionality, which requires large spans. The authors managed to solve this by introducing projection pursuit to convert multi-variables into a uni-variable. The idea is to estimate multivariate functions by combinations of carefully selected linear combinations of smooth univariate functions. We need an initial model, and then update the model to narrow down the difference between the model and the sample distribution. Then the authors gave a detailed mathematic proof of the ration of data and the model marginals along a certain direction. They also discussed about some important details about redundant-variables and stop criterion. In the experiment part, the author compared their method with the k-nearest neighbor method and shows robustness to high-dimension noise.

Strong points

1. **Originality.** The author successfully expanded the problem from unvaried cases to multivariate cases. Previous work all failed to give a low-biased multivariable estimation. This paper, by introducing projection pursuit, successfully give a novel solution to this problem.

2. **Clarity.** The paper is well written and easy to understand. The authors organized the paper in a clear way. From theory to technical details and then the experiments, this paper give us a step-to-step guidance into this method. They also included adequate information for further reference.

3. **Quality.** The authors gave a detailed mathematic proof of the ration of data and the model marginals. This makes the paper technically sound. The authors also gave a concrete procedure to avoid redundant-variables and to decide stop criterion. This increase the reproducibility of this method.

Weak points

1. **Experiments.** The first two experiments are most from parameterized Gaussian model, which is hard to encounter in real life. It is also not a non-parameterized distribution as claimed in the previous theory. Using a combination of Gaussian distribution to estimate itself is meaningless, since we are doing estimation based on the priori. The distribution will not be Gaussian in most real case, achieving a high accuracy under this circumstance will be more convincible.
Also the size of dataset is relatively small. All of the three experiments use less than 1000 data points. The author also does not discuss how the amount of data will affect the model accuracy, which is of great importance. We also have to know the expected data amount before training and deciding when to not collect more data.
For the last experiment, the author just decide the model to be a product of two dimensional marginal densities according to factored approximation. The factored approximation features are often difficult to compute for high-dimensional dataset. They also did not compare this method with other methods such as k-nearest-neighbor in this experiment, which is less convincing.

2. **Technical details.** On critical point is that we have to assign the initial model and the initial probability value. The author think it is an advantage. However, priori information is often difficult to acquire especially when we know little about target dataset. The author said a Gaussian density is often a natural choice but also not give a concrete explanation.
The author did not give enough reason why we should use a product of univariate functions as the model. In general, the authors did not tell us how to decide the model density function and their initial value.
The author said we can add graphical information to help gain insight toro the data distribution. However, for high dimension data, it is often hard to construct graphical information. Further more, using graphical information to help machine learning is like using human judgement and experienced knowledge to help training. This does not give a solution to solving general problems.