Application of Adaptive Mixtures and Fractal Dimension Analysis Technique to Particle Physics

S.J. Lee, P. Padley, B. Chase
Rice University, Houston, TX 77005, U.S.A.

In this paper, we examine the applicability of using Adaptive Mixtures to represent high energy physics data. We attempt to use Adaptive Mixtures to derive a discriminant variable to identify tau leptons in hadron collider data. In addition, we examine the applicability of Fractal Dimensions as a tool to search for physics signals.

1. INTRODUCTION

The discrimination of physics "signal" from "background" is one of the most important subjects in high energy physics analysis since this process usually governs the magnitude of measurement errors. Background suppression using kernel density estimation to estimate the parent distribution of a data sample appears to be an effective method. In this paper, Adaptive Mixtures [1] and Kernel Density Estimation with Likelihood Maximization (KDELM) are examined. In addition, the Fractal Dimension [2] of a data set is studied as a tool to find physics signals.

Adaptive Mixtures are designed to accept the strengths of the kernel estimates and that of the finite mixtures while discarding the weaknesses of these methods. In the kernel estimation method, one kernel is assigned to each datum. As a result, kernel density estimates converge asymptotically under very weak conditions. Due to the large number of kernels, however, there is computational disadvantage in the case that there are many data points. On the other hand, since finite mixture methods employ small fixed numbers of kernels, these approaches have an advantage in calculation time. However, very strong assumptions must be put on the underlying densities as well as on the initial states of kernels. In the adaptive mixture, the number of kernels is driven by data. At each data point, criteria for making the decision whether a new kernel is added are tested. When these criteria are satisfied, a new kernel is added. Otherwise, the parameters of the adaptive mixture kernels are updated to accommodate the new data point. In this way, only the necessary number of kernels are introduced in the fit while the characteristic of asymptotic convergence remains without strong assumptions on the underlying distributions.

KDELM uses the maximization of likelihood. The main difference between KDELM and Adaptive Mixtures is that, in KDELM, all data points are taken into account at the same time to calculate the likelihood of the kernel estimates. Only when addition of a kernel reduces the overall likelihood is the new kernel added to the kernel estimates. If the new kernel produces a worse overall likelihood, the process adding new kernels is ceased.

The fractal dimension [2] D, also called capacity dimension, is defined by

$$n(\epsilon) = \epsilon^{-D}, \qquad (1)$$

where $n(\epsilon)$ is the minimum number of open sets of diameter ϵ a to cover the set. Fractal dimension quantifies the increase in structural definition that magnification yields. As an example, a good estimation of the length of a coast line can be obtained using a meter stick. However, if a centimeter stick is used in this measurement, a better and larger estimation of the coast line length will be obtained. Fractal dimension quantifies this increase in detail that occurs by magnifying or, in this case, by switching rulers. The motivation to use the fractal dimension in the signal discrimination process is that the fractal dimension would efficiently identify signal from background if the geometrical configurations of signal and background in the d-dimensional space are significantly different. Furthermore, since the calculation of fractal dimension does not contain a complicated logical algorithm, the fractal dimension calculation has an advantage in calculation time.

2. METHODS

2.1. Adaptive Mixtures

The kernel density estimate K consists of N_K kernels, and each kernel K_i has a weight w_i. In d-dimensional space, a kernel K_i is a multivariate product of d univariate Gaussians. As a result, K_i is characterized by d-dimensional mean and variance, μ_i and σ_i, respectively. The kernel density estimate K at a d-dimensional data point x is represented by

$$K(x) = \sum_{i=1}^{N_K} w_i K_i(x; \mu_i, \sigma_i). \qquad (2)$$

In Adaptive Mixtures, the number of kernels is driven by data. The kernel density estimate K^n found by the data points $x_1, x_2, ..., x_n$ is given by

$$K^n = \sum_{i=1}^{N_K^n} w_i^n K_i(\mu_i^n, \sigma_i^n). \qquad (3)$$

A new kernel is added at the next data point x_{n+1} when the Mahalanobis distance from the new data point to each kernel is greater than a predefined threshold T_c. If this criterion is not satisfied, then w_i^n, μ_i^n and σ_i^n are updated to w_i^{n+1}, μ_i^{n+1} and σ_i^{n+1} without addition of a new kernel. The update and creation rules are specified in reference [1]. The Mahalanobis distance M between a one-dimensional data point x and a kernel with mean μ and standard deviation σ is defined by

$$M = \frac{(x-\mu)^2}{\sigma^2}. \tag{4}$$

In this study, the threshold T_c is set to be 9.

2.2. Kernel Density Estimation with Likelihood Maximization

In this method, addition of a new kernel takes place only when the addition produces better fit result. The goodness of fit is estimated by examining the log-likelihood given by

$$\log\mathcal{L} = \sum_i K(x_i), \tag{5}$$

where the sum runs over data points and the kernel estimate $K(x_i)$ is defined in equation 2. The constraint that the kernel weights sums to one is applied.

2.3. Fractal Dimension Calculation

There are several techniques to calculate fractal dimension, yet all involve estimating the dimension from the slope of a log-log power law point. The technique used in this study is the box counting technique [2]. The specific process for the fractal dimension calculation is:

1. Grids or boxes of varying side lengths are placed over data sample.

2. A count of how many boxes contain data points is made for the power low plot.

3. From the least square fit of the slope of the power low plot, calculate the fractal dimension.

3. RESULTS AND CONCLUSIONS

3.1. Adaptive Mixtures

To test the performance of the adaptive mixture method, two one-dimensional data samples, which are randomly derived from $(1/\sqrt{2\pi})\exp(-x^2)$ and $\exp(-x)$ for $x \geq 0$, are used. Each data sample contains 100 data points and the fit results obtained using Adaptive Mixtures is shown in Fig. 1.

Table I ϵ_S/ϵ_B comparison in various statistical tools

Method	ϵ_S/ϵ_B
KDELM	26.32 ± 4.51
Decision Tree [3, 4]	6.07 ± 1.24
Neural Network [4, 5]	5.98 ± 1.20
Support Vector [4, 6]	3.95 ± 0.42

In the example of the Gaussian distribution shown in the left plot of figure 1, the data is over-fitted. Furthermore, the fit in the exponential example given in figure 1 is not consistent with the data points. It is found that these problems cannot be resolved by changing the value of the threshold T_c. As a result, a new algorithm may be necessary to obtain a better iteration result and to prevent over-fit.

3.2. Kernel Density Estimation with Likelihood Maximization

KDELM is considered as a solution to fix the problems with Adaptive Mixtures. The performance of KDELM for the Gaussian and exponential distributions is quite satisfactory, which is shown in the figure 2.

As another test, this technique is applied to tau lepton identification. A Monte Carlo (MC) sample for the decay mode $W^- \to \tau^- \nu_\tau$ (charge conjugation symmetry is assumed) generated for a hadron collider experiment and a generic Quantum Chromodynamics MC sample are used as "signal" and "background", respectively. A discriminant function defined by

$$D(x) = \frac{K_S(x)}{K_S(x) + K_B(x)}, \tag{6}$$

is considered to separate signal MC events from background. Here, $K_S(x)$ and $K_B(x)$ are the signal and background kernel estimates, respectively. Figure 3 shows the distribution of $D(x)$. To compare the signal discrimination power of various statistical tools, ϵ_S/ϵ_B's obtained using these tools are compared, where ϵ_S is the probability for a signal event to be identified as signal and ϵ_B is the probability for a background event to be identified as signal. In this ϵ_S/ϵ_B comparison, ϵ_S is fixed to be 50% and the results are shown in table I.

From these two tests, KDELM turns out to be (i) very robust in the fit, (ii) fast in computation, and (iii) good in signal discrimination. In the signal discrimination, KDELM is as good as the neural network technique, while with KDELM it is conceptually easier to understand the whole fit process.

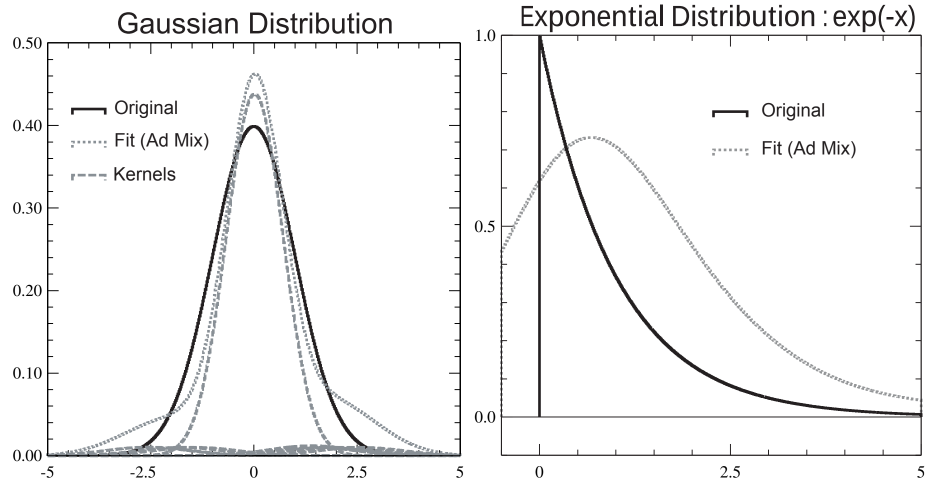

Figure 1: Fit results obtained using Adaptive Mixtures for data samples generated from (left)$(1/\sqrt{2\pi})\exp(-x^2)$ and (right) $\exp(-x)$ for $x \geq 0$.

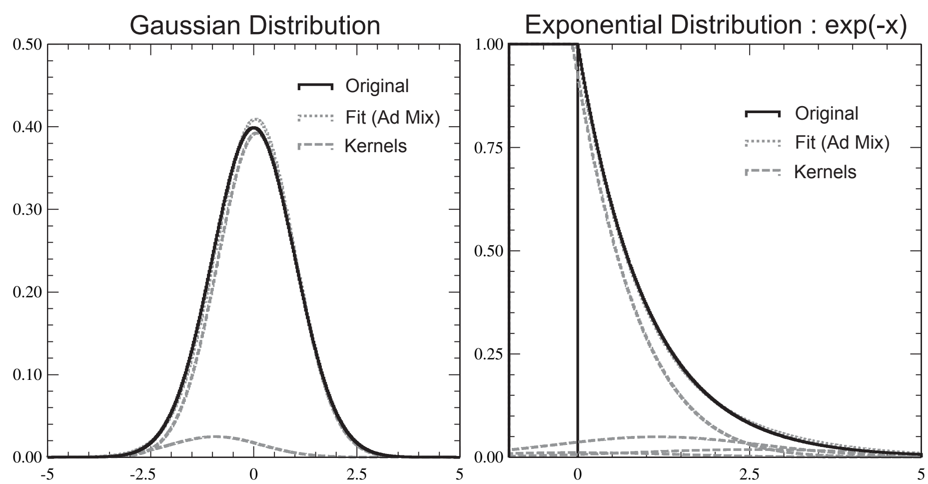

Figure 2: Fit results obtained using KDELM method for data samples derived from (right)$(1/\sqrt{2\pi})\exp(-x^2)$ and (left) $\exp(-x)$ for $x \geq 0$.

3.3. Fractal Dimension Calculation

To check whether fractal dimension calculation will be useful in high energy physics analysis, we use the same MC samples used in the second test of KDELM. First, using the fractal dimension technique, we find the combinations of physics variables which provide the best signal discrimination. Then, these variable combinations are compared with those found using KDELM. Due to lack of statistics of the MC samples, this comparison is taken only up to two-variable combinations. However, this comparison reveals that the best discriminant variables found using fractal dimension often disagree with those obtained using KDELM. As a result, the applicability of fractal dimension calculation as a discriminating feature in high energy physics analysis is skeptical.

Acknowledgments

The authors wish to thank professor David Scott (Rice University) and professor Bruce Knuteson (MIT) for their valuable discussion. Also, the authors would like to thank to professor Ricardo Vilalta (University of Houston) and Abraham Bagher for their calculation of ϵ_S/ϵ_B's in various statistical techniques other than KDELM.

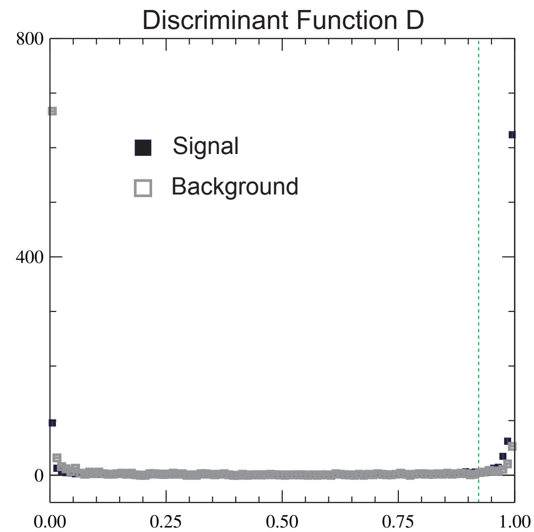

Figure 3: The distribution of $D(x)$ in the tau lepton identification study. The dashed line shows the location of $D_{1/2}(x)$. When we select an event as signal if its $D(x)$ is greater than $D_{1/2}(x)$, ϵ_S becomes 50%.

References

[1] C.E. Priebe, "Adaptive Mixtures", *Journal of the American Statistical Association*, **89**, 796 (1994).

[2] S.N. Rasband, "Fractal Dimension" Wiley, 1990.

[3] S. Russell and P. Norvig, "Artificial Intelligence: A modern approach", Prentice Hall, 1995.

[4] T.M. Mitchell, "Machine Learning", McGraw Hill, 1997.

[5] J. Hertz *et al.*, "Introduction to the Theory of Neural Computation", Addison-Wesley, 1991.

[6] N. Cristianini and J. Shawe-Taylor, "An Introduction to Support Vector Machines and Other Kernel-based Learning Methods", Cambridge University Press, 2000.