

Paper of Interest

R. Kondor and T. Jebara. A Kernel Between Sets of Vectors. In Proceedings of International Conference on Machine Learning, Washington, D.C., Aug 2003.

<http://people.cs.uchicago.edu/~risi/papers/KondorJebaraICML03.pdf>

My Tasks

Generally speaking, I am going to implement two probability product kernel instead of the base kernel, one is implemented on the sample space, the other is implemented on the Hilbert Space. This includes several process:

1. Understanding the Hilbert space, Gaussian Process, Kernel PCA
2. Implementing the Gaussian RBF directly on the sample space
3. Map the vectors to a Hilbert space
4. Fit different Gaussian distributions to the whole set using Kernel PCA
5. Calculate kernels between sets using Bhattacharyya's measure of affinity between Gaussians

Relations to My Work

My recent research involves visualizing high dimensional earthquake simulation data. After applying a splitting methods, each earthquake simulation becomes a set of vectors. Using base kernel methods, we can calculate the kernels between different set of vectors. But the base kernel is limited and may not capture all the features. Instead, Gaussians yields interesting nonlinear kernels which will describe the data in a better way.

Expectations

1. Given the current dataset, complete the Gaussian RBF on the sample space
2. Complete the mapping from the vector dimension to the Hilbert space
3. Complete the fitting Gaussian distributions for different sets of vectors Using Kernel PCA
4. Complete the kernel matrix between different set of vectors
5. Visualizing the kernels using adjacency matrix, bar charts and other plots.

Evaluation

My evaluation includes two parts:

1. From theoretical perspective, I would compare the probability product kernels with the base kernels and explain why it provides a better way of explaining data
2. A practical comparison between two different kernel methods, from the visualization, we can directly observe the data and give an evaluation about which kernel method works better.