Kernel Approximation via Empirical Orthogonal Decomposition for Unsupervised Feature Learning

Yusuke Mukuta
The University of Tokyo
7-3-1 Hongo Bunkyo-ku, Tokyo, Japan
mukuta@mi.t.u-tokyo.ac.jp

Tatsuya Harada
The University of Tokyo
7-3-1 Hongo Bunkyo-ku, Tokyo, Japan
harada@mi.t.u-tokyo.ac.jp

A. Feature functions assuming Gaussian distribution

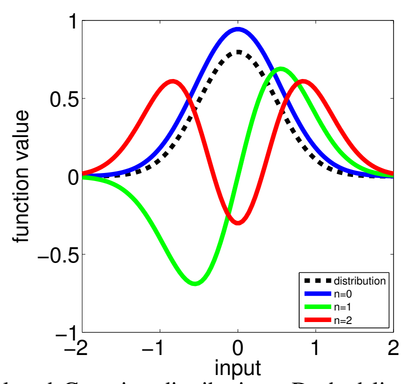

Figure 1. Eigenfunctions for a Gaussian kernel and Gaussian distribution. Dashed line represents the distribution and the solid lines represent the feature functions.

B. List of architectures used in the experiments of unsupervised feature learning

Setting	Input	N	m_1	p_1	γ_1	m_2	p_2	γ_2
MNIST								
1	Patch	2	5×5	50	2	2×2	200	2
2	Grad	2	1×1	12	2	3×3	50	2
3	Grad	2	1×1	12	2	3×3	400	4
CIFAR-10, CIFAR-100, SVHN								
1	Grad	2	1×1	12	2	2×2	800	4
2	Patch	2	2×2	100	2	2×2	800	4
3	Patch	1	3×3	800	10	–	–	–

Table 1. List of architectures reported in the paper. With regard to the Input column, "Patch" indicates that the network works on patches. "Grad" indicates that the network works on gradient maps. N represents the number of layers. p_i represent the number of filters for i-th layer; m_i represent the size of the patches; γ_i is the subsampling factor.