A Deep Neural Network Based Quasi-Linear Kernel for Support Vector Machines

Weite LI and Jinglu HU
Graduate School of Information Production and Systems, WASEDA University

1. SVM with a Quasi-Linear Kernel

$$f_P(x)=\sum_{j=1}^{M}(\Omega_j^T x+b_j)R_j(x)+b$$

$$\Big\Downarrow \quad \Phi(x)=\left[R_1(x), x^T R_1(x), \ldots, R_M(x), x^T R_M(x)\right]^T, \quad \Theta=[b_1,\Omega_1^T,\ldots,b_M,\Omega_M^T]^T$$

$$f_P(x)=\Theta^T\Phi(x)+b$$

$$\Big\Downarrow$$

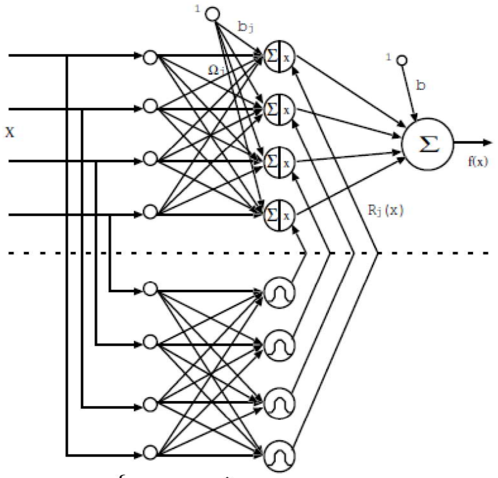
$$\max_{\alpha}\ \mathcal{J}_D(\alpha)=-\frac{1}{2}\sum_{k,l=1}^{N} y_k y_l K(x_k,x_l)\alpha_k\alpha_l+\sum_{k=1}^{N}\alpha_k$$

$$s.t.\begin{cases}\sum_{k=1}^{N}\alpha_k y_k=0\\ 0\le\alpha_k\le c, k=1,\ldots,N\end{cases}$$

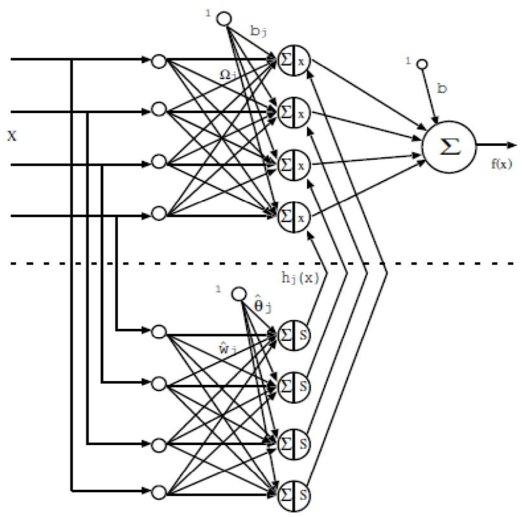
$$K(x_k,x_l)=\Phi^T(x_k)\Phi(x_l)=(1+x_k^T x_l)\sum_{j=1}^{M}R_j(x_k)R_j(x_l)$$

2. A Network View

- When $R_j(x)=\exp\left(-\|x-\mu_j\|_F^2/\lambda\sigma_j^2\right)$, it provides a flexible and adjustable kernel, filling the gap between linear and nonlinear kernels.

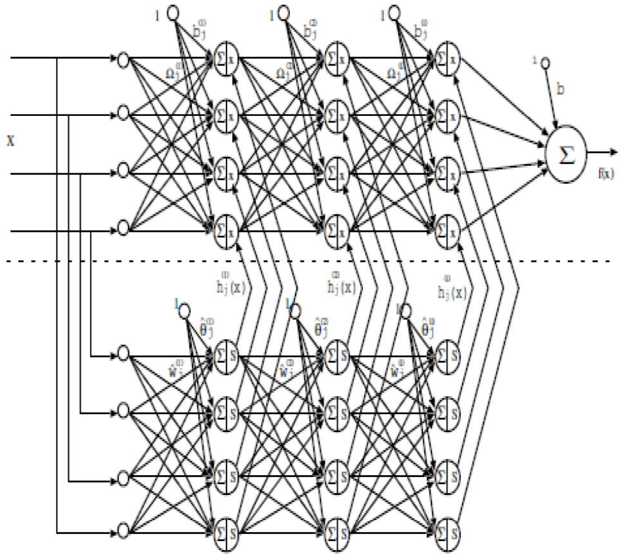


- When $R_j(x)=\begin{cases}1, & \hat{w}_j^T x+\hat{\theta}_j>0\\ 0, & \hat{w}_j^T x+\hat{\theta}_j\le 0\end{cases}$, it mimics the functionality of a rectified network with single hidden unit, whose multilayer versions are widely used.

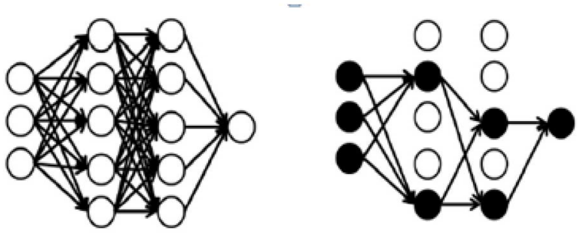


3. An Extension to Multilayer Version

- When multiple layers stacked together, a deep quasi-linear kernel can be built upon it.



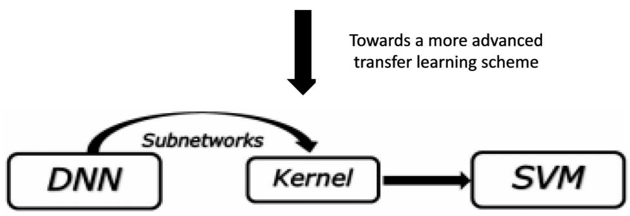
- It uses the subnetworks of a pre-trained deep neural network to built a kernel for SVM. A subnetwork looks like the followings, which only comprises a subset of parameters.



(a) The framework of a neural network (b) A subnetwork for an input object

- It leads to a more advanced transfer learning scheme, which makes the tuning of deep neural networks also available for small data sets without the sacrifice of generalization ability.

(a) The transfer of features



(b) The transfer of subnetworks

古月研究室
早稲田大学大学院情報生産システム研究科
Email: {liweite@akane; jinglu@}waseda.jp.
Tel/Fax: 093-692-5271