A Convenient Kernel

Target

Construct a discriminant hypersurface, which separates the positive ones from the negative ones.

Training data set

$$D=\{(x_i,\ y_i): x_i \in X \subset R^{d_x},\ y_i=-1,\ or\ 1\ \}_{i=1}^{l}$$

Approach --- Kernel plug-in

Map the training data in the input space X into a higher-dimensional feature space Z via a transformation ϕ

- Transformation

$$\phi: X \to Z$$

$$z=\Phi(x)=(\sqrt{\lambda_1}\phi_1(x),\ \dots,\ \sqrt{\lambda_{d_z}}\phi_{d_z}(x))^T \quad (d_z \le \infty)$$

where $\{\phi_k\}_{k=1}^{d_z}$ are linarly independent

and $\lambda_1 \ge \lambda_2 \ge \dots \ge \lambda_{d_z} > 0$

- Kernel

$$K(x,\ u)=\sum_{k=1}^{d_z}\lambda_k\phi_k(x)\phi_k(u)$$

- Discriminant function

A test point x is classified according to the sign of the following discriminant function

Fisher Discriminant

- Other kernels

(1) The classical Fisher linear discriminant (FLD) can be represented by the proposed reproducing kernel Hilbert space (RKHS) framework. The corresponding kernel is given by $K(x,\ u)=x'\Sigma^{-1}u$, where Σ is the pooled within-class covariance matrix.

(2) Regularized discriminant analyses can reproducing kernel of form $K(x,\ u)=x'M^{-1}u$, where M is a positive definite matrix.

Simulation study

- Training sample were generated from

$$x_j \sim Uniform\{(-1,1)^2\}$$

- Corresponding class labels were generated from

$$y_j=sign\{x_{j2}-\sin(\pi x_{j1})+\varepsilon_j\},$$

$$\varepsilon \sim N(0,\sigma^2),\ \sigma^2=0.01$$

- Discriminant function

$$f(x)=x_2-\sin(\pi x_1),\quad x=(x_1,\ x_2)\in[-1,\ 1]^2$$

- Kernel: radial basis function (rbf) kernel with $\tau=0.1$

$$K(x,u)=\exp\{-\|x-u\|^2/(2\tau^2)\}$$

- Training data of size: 300; testing data of size: 200
- Accuracy: 97.00 %

$$\hat{f}(x) = \|z(x) - \bar{z}_{-1}\|^2_{l_2(z)} - \|z(x) - \bar{z}_1\|^2_{l_2(z)}$$

$$= \frac{-2\sum_{j\in I_{-1}} z' z_j}{l_{-1}} + \frac{\sum_{j,k\in I_{-1}} z'_j z_k}{l_{-1}^2} - \frac{-2\sum_{j\in I_1} z' z_j}{l_1} + \frac{\sum_{j,k\in I_1} z'_j z_k}{l_1^2}$$

$$= \frac{-2\sum_{j\in I_{-1}} K(x, x_j)}{l_{-1}} + \frac{\sum_{j,k\in I_{-1}} K(x_j, x_k)}{l_{-1}^2}$$

$$- \frac{-2\sum_{j\in I_1} K(x, x_j)}{l_1} + \frac{\sum_{j,k\in I_1} K(x_j, x_k)}{l_1^2}$$

where $z_j = \Phi(x_j)$ and $\bar{z}_i = \sum_{j\in I_i} \frac{\Phi(x_j)}{l_i}$

Expect $\begin{cases} \hat{f}(x) > 0 & \text{if } y=1 \\ \hat{f}(x) < 0 & \text{if } y=-1 \end{cases}$

Real applications

- **Datasets:**
  - (1) Wisconsin Diagnostic Breast Cancer (WDBC)
  - (2) Ionosphere
  - (3) Liver-disorders
- **Data information:**
  - (a) Number of instances
  - (b) Number of attributes
  - (c) Class distribution
  - (d) Tuning parameters: RBF kernel with τ
  - (e) 10-fold cross validation with accuracy
  - (f) Average accuracy

Kernels

- Uniform kernel

$$f(x) = \frac{n_{1x}}{l_1} - \frac{n_{-1x}}{l_{-1}} + b$$

$$b = \frac{1}{2}\left(\frac{n_{-1}}{l_{-1}^2} - \frac{n_1}{l_1^2}\right)$$

where n_{1x} (n_{-1x}) is the number of positive (negative) training inputs and pairs, respectively. Their distance is less than the kernel window width h.

Info. \ Data	WDBC	Ionosphere	Liver-disorders
a	569	351	345
b	30	31	6
c	357 benign, 212 malignant	225 good, 126 bad	145 for group1, 200 for group 2
d	$\tau = 5$	$\tau = 1$	$\tau = 0.5$
e	.89 .88 .93 .93 .95 .98 .89 .98 .95 .93	.94 .97 .93 .93 .93 .94 .97 .99 .97 .97	.57 .60 .69 .51 .57 .76 .71 .74 .59 .59
f	93.16 %	95.40 %	63.25 %