Course: Math 6397 Spring 2015
Instructor: Robert Azencott
Lecture 3 Date: 1/27/2015
Notes prepared by Christopher Cote
Draft 2

Reading Assignment:

Read "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", Ch.4 Linear Discriminant Analysis pgs. 84 - 92. Written By Hastie, Tibshirani, Friedman.

1^{st} Project:

Data Sets → UCI Machine Learning Repository: Data Sets
Website: https://archive.ics.uci.edu/ml/datasets.html

- Look at 2-3 data sets.
- Formulate a description in terms of
  - number of classes
  - number of cases
  - number of attributes for description of each "case"

Work to perform on data sets: (Using R, Matlab, or other scientific programming language)

1. Present histograms of distributions for the value of typical attributes.
   - Present Range of values, mean value of attribute, standard deviation of attribute.
2. Pick two classes:
   - generate linear discriminate analysis classification for your 2 chosen classes.
   - Complete performance evaluation of this classifier.

1 Kernel Methods

The function $K(s,t)$ is defined on $S \times S$ where:

- s, t are "objects"

- $s, t \in \mathcal{S}$ of objects
- $s \to v(s) \in R$
- K is a positive definite kernel

Important rules to quantify similarity by a kernel:

1. $K(s,t) = K(t,s)$
2. $s_1, ..., s_N \in S$; (N is arbitrary)
3. $M_{ij} = K(s_i, s_j)$; $M = M_{ij}$
4. $N \times N$ matrix must be semi-positive definite

Examples of Kernels:

Case where $S \subset R^k$; $s \to v(s); v(s)$ is the vector of attributes of s and belongs to R^k; $K(v,w); v, w \in R^k$

1^{st} Example

$$K(v,w) \;= <v,w> = \; v^t w;\; v, w \in R^k$$

K is a positive definite kernel.

2^{nd} Example

Polynomial Kernel

$$[a+ <v,w>]^r \;= [1+ <v,w>]^r \;= (1 + v_1 w_1 + v_2 w_2 + ... + v_k w_k) = K(v,w)$$

where a is constant,$v, w \in R^k$, r is a fixed integer. K is a kernel $\forall a$, $\forall r \geq 2$

3^{rd} Example

$$K(v,w) = exp\left[-\frac{||v-w||^2}{\sigma^2}\right];\; \sigma > 0$$

$||v||^2 = v^* v = <v,v>$. This is a positive definite kernel. $K(v,v) = 1$.

$v, w \in R^k$

Generic Kernels:

Hilbert Space (H) is an abstract vector space endowed with a scalar product. $<v,w>_H \in R$.

- $v, w \in H,\; v + w \in H,\; \lambda v \in H,\; \lambda \in R$

Properties of Hilbert Space:

- $<v,w> = <w,v>$ (symmetry)

- $< v_1 + v_2, w >_H = < v_1, w >_H + < v_2, w >_H$ (linearity)
- $< v, v >_H \geq 0$ and $< v, v > = 0$ is equivalent to $x = 0$ (positive definite)
- $||v||^2 = < v, v >$
- $d^2(v, w) = < v - w, v - w > = ||v - w||^2$
- H with distance d is a metric space (X,d) and is said to be complete if every Cauchy sequence in X converges to a point in X.

Examples of Hilbert Space:

1. $H = L^2(R^3); f(x) \in R, \ x \in R^3$

$$\int_R |f(x)|^2 dx < \infty == < f, g >_H = ||f||^2$$

2. $f(\theta): \ \theta \in (0, 2\pi)$ (periodic function of θ)

$$\int_0^{2\pi} \Big[f(\theta)\Big]^2 d\theta < \infty; \ L^2([0, 2\pi]) = Hilbert\ Space$$

3. Fourier Series of f

   Note: the sine and cosine functions are orthogonal in H and have the same norm.

$$c_j = \int_0^{2\pi} f(\theta)\ cosj\theta\ d\theta$$

$$f(\theta) = c_0 + c_1 \cos\theta + s_1 \sin\theta + c_2 \cos 2\theta + s_2 \sin 2\theta$$

Note: Many spaces of functions of interest in functional analysis are Hilbert Spaces.

Generic Kernel:

S = set of objects ; $s \in S \rightarrow \ \Phi(s) \in H \ ; \ \Phi : S \rightarrow H$

Define a kernel: $K(s, t) = < \Phi(s), \Phi(t) >_H$

Theorem: If K is a positive definite kernel $K(s, t)$ defined on S, and there is a Hilbert Space H, then there is a function $\Phi : S \rightarrow H$ such that $K(s, t) = < \Phi(s), \Phi(t) >_H$.

Radial Basis Function Example:

$$K(s, t) = exp\left[-\frac{||s - t||^2}{\sigma^2}\right]; \ s, t \in R^k; \ S = R^k$$

H is the smallest Hilbert space containing the space of all finite linear combinations of radial basis function.

$\Phi(s)$ is a radial basis function if there is a vector $w \in R^k$ such that:

$$\Phi(s)_w = exp\left[-\frac{||s-w||^2}{\sigma^2}\right];$$

σ is fixed; $\forall s \in R^k$

$a_1\Phi_{w_1}(s) + a_2\Phi_{w_2}(s) + ... + a_j\Phi_{w_j}(s)$ (linear combination of radial basis functions);

Scalar product is defined by the kernel:
$K(w_1, w_2) \ = <\Phi_{w_1}, \Phi_{w_2}> = exp(-||w_1 - w_2||^2)$
$= exp(-(w_{1_1} - w_{2_2})^2 - (w_{2_1} - w_{2_2})^2$
$= exp(-w_{1_1}^2 + 2w_{1_1}w_{2_2} - w_{1_2}^2 - w_{2_1}^2 + 2w_{2_1}w_{2_2} - w_{2_2}^2)$
$= exp(-||w_1||^2)\exp(-||w_2||^2)\exp(2w_1^T w_2)$