Classnote 3 Automatic Learning and Data Mining

Peixin Zhang 0891164

February 6, 2015

1. Let S be a set of n objects, and $s, t \in S$. A function $k : S \times S \to \mathbb{R}$ is called a positive definite kernel, provided $k(s,t) \in \mathbb{R}$, $k(s,t) = k(t,s)$ and

$$\sum_{i=1}^{n} \sum_{i=1}^{n} c_i c_j k(s_i, t_j) \geq 0$$

for arbitrary $n > 0$, any choice of n objects $s_i, \cdots, t_j \in S$, and any arbitrary value of $c_i, c_j \in \mathbb{R}$. For $s_1, \cdots, t_N \in S$, define $M_{ij} := k(s_i, s_j), M = (M_{ij})_{N \times N}$, we want M to be semi positive definite matrix which means the eigenvalues of M should not be negative. Following, we will show three groups of kernels, including: linear kernel, polynomial kernel and Gaussian kernel, which can be used to define matrix M..

2. Use following Thm to define a positive definite kernel

Thm: Any positive definite kernel k on a space S can be defined by constructing as

$$k(s,t) =< \phi(s), \phi(t) >$$

from a self reproducing Hilbert space $\mathbb{H}$ and a mapping $\phi : S \to \mathbb{H}$ with $\phi(s) \in \mathbb{H}$.

Example: For $x = (x_1, x_2, \ldots, x_k) \in \mathbb{R}^k$, $\phi(x) = v = (v_1, v_2, \ldots, v_k, \ldots) \in \mathbb{H}$ satisfies $\sum_{i=1}^{\infty} v_i < \infty$.

3. Hilbert space $\mathbb{H}$ construction

3.1 We first define a self reproducing Hilbert space $\mathbb{H}$ which is a vector space of functions on S, but not all functions defined on S are belongs to $\mathbb{H}$. We define functions with the following type: $s \in S$,

$$\phi_s(t) = k(s,t), \forall t \in S$$

which are functions of similarity measurement.

3.2 Let N be an arbitrary number. Take N arbitrary objects $(s_1, s_2, \cdots, s_N) \in S$, and define N functions $\phi(s_1), \phi(s_2), \cdots, \phi(s_N)$ as above. Then by considering finite linear combination of these N functions $\phi(s_1)\lambda_1 + \phi(s_2)\lambda_2 + \cdots + \phi(s_N)\lambda_N$ with $\lambda_i \in \mathbb{R}$, we obtain a vector space of all linear combinations of ϕ_s.

3.3 Based on our vector space, we define a scaler product on it to have a Pre-Hilbert space.

Know kernel k is positive definite. Suppose

$$\phi_s = \sum_{j=1}^{N} \lambda_j \phi_{s_j} and \phi_t = \sum_{k=1}^{R} \mu_k \phi_{t_k}$$

Define

$$k(s,t) =< \phi_s, \phi_t >,$$

then $< \phi_s, \phi_t >= \sum_{j=1}^{N} \sum_{k=1}^{R} \lambda_j \mu_k < \phi_{s_j}, \phi_{t_k} >= \sum_{j=1}^{N} \sum_{k=1}^{R} \lambda_j \mu_k < s_j, t_k > \geq 0$

and if $s = t$, $k(s,t) =< \phi_s, \phi_t >=< \phi_s, \phi_s >= \|s\|^2 = \|\sum_{j=1}^{N} \lambda_j \phi_{s_j}\|^2$.

3.4 Last, let V be the vector space of finite linear combination of functions $\phi_x \in \mathbb{H}$ equipped with scaler product $k(s,t) =< \phi_s, \phi_t >$. By taking completion of this pre-Hilbert space, we define a self-producing Hilbert space. And this Hilbert space is the smallest space containing V.

- Remark 1:Completion means the following: Give $\phi_s = \lambda_1 \phi_1 + \lambda_2 \phi_2 + \cdots + \lambda_k \phi_k + \cdots$ with $(\|\phi_s\|)^2 =< \phi_s, \phi_s >$, we have $\sum_{k=1}^{\infty} (\lambda_k)^2 < \phi_s, \phi_s >< \infty$.
- Remark 2:We could also write $\phi_s(t) = k(s,t), \forall t$ as $\phi_s = k(s, \cdot)$, thus $< k(s, \cdot), k(t, \cdot) >_{\mathcal{H}} = k < s, t >$.
- Remark 3: Kernel are often presented as measure of similarity. Function $v \to \phi_v = k(v,t), \forall t$ is a list of similarity between v and the other objects. Often, we know how to define similarity between two objects.
- Remark 4: Hilbert space $\mathbb{H}$ is a space of linear combination of similarity functions for each objects.

4. Hilbert space construction examples

Assume a object $s \in \mathbb{R}^p$ with p attributes, and s is defined by $v(s) = (v_1, v_2, \cdots, v_p)$. Kernel k is defined as $k(s,t) = k(v(s), v(t))$, following are three kernel types

4.1 Linear kernel

$$k(v,w) =< v, w >_{\mathbb{R}^k} = v_1 w_1 + v_1 w_1 + \cdots + v_p w_p$$
$$k(v,v) = \|v\|^2 = (v_1)^2 + (v_2)^2 + \cdots + (v_p)^2$$

Take $v \in S \sim \mathbb{R}^p$, define $\phi_v(t) = k(v,t) =< v, t >$ for $\forall t \in S$, then arbitrary N linear combination

$$\psi(t) = \sum_{j=1}^{N} \lambda_j \phi_{v_j}$$

implies $\psi(t) =< u, t >_{\mathbb{R}^p} \in \mathbb{H}$ where $u = \sum_{j=1}^{N} \lambda_j v_j$ and this self producing Hilbert space is the dual space of $\mathbb{R}^p$ with scaler product

$$< \psi(v), \psi(w) >= k(v,w)$$

.

4.2 Polynomial kernel

$S = \mathbb{R}^p$. Let $k(v,w) = (1+ < v, w >_{\mathbb{R}^p})^2$. Fixed $v \in S$, define $\phi_v(t) = (1+ < v, t >)^2$ for $\forall t \in S$.

Example: $p = 3$, thus $\phi_v(t) = (1 + v_1t_1 + v_2t_2 + v_3t_3)^2$ which is a polynomial of degree two w.r.t t.

For arbitrary large N, take arbitrary objects $v_1, v_2, \cdots, v_N \subseteq S$, the finite linear combination

$$\lambda_1\phi_{v_1} + \lambda_2\phi_{v_2} + \cdots + \lambda_N\phi_{v_N} \in \mathbb{H}$$

is a polynomial of degree 2 w.r.t. t with dimension 10 when assume $p = 2$. Thus $\mathbb{H}$ is a space of all polynomial with degree of 2.

In $\mathbb{H}$, scale product is defined as

$$< \phi_v(t), \phi_w(t) >_{\mathbb{H}} = (1+ < v, w >)^2$$

and

$$< \phi_v(t), \phi_v(t) >_{\mathbb{H}} = (1+ < v, v >_{\mathbb{R}^3})^2 = \|v\|^2$$

Generally speaking, if $k(v,w) = (1+ < v, w >_{\mathbb{R}^k})^R$, then $\mathbb{H}$ is a Hilbert space with polynomial of degree R.

4.3 Gussian kernel

Define kernel k as

$$k(v,w) = \exp(\frac{-\|v-w\|^2}{\sigma^2})$$

Fixed $v \in S$, define

$$\phi_v(t) = \exp(\frac{-\|v-t\|^2}{\sigma^2})$$

for $\forall t \in S$, which is a radial function,then $\mathbb{H}$ is an infinite dimension complete space generated by linear combination of radial functions, i.e. series of radial functions which are bounded equiped with scaler product

$$< \phi_v, \phi_w >_{\mathbb{H}} = exp(\frac{-\|v-w\|^2}{\sigma^2})$$

.

Remark: mapping ϕ replace finite dimension vector with infinite dimension vector in $\mathbb{H}$.