1 背景知识

在我们详细介绍算法细节之前，让我们先了解一下关于图分析计算的所需的基本背景。这章包含 (i) 图核函数，(ii) 图小波分析两部分。

1.1 图

一个*标号图*可以被一个有限的节点集合 V 和一个有限的边集合 $E \in V \times V$ 所描述。在大多数应用中，图都是有标号的。这些标号都是从一个标号集选取的，我们用一个标号函数 $\lambda : V \cup E \to \Sigma$ 来给各个节点和边分配标号。在*标号点图*中只有点有标号，同样，在*标号边图*中只有边有标号。在*全标号图*中，边点都有标号。如果用一种特殊的标号来表示未标号的点和边，那么标号点图和标号边图都可以被看做是全标号图的特殊形式。所以在此文中我们只考虑全标号图来简化问题而又不失一般性。对于标号集 Σ 我们并不指定具体结构，可以是一个字段, 一个向量，也可以是很简单的是个集合。以下我们约定，一幅图用一个四元组 $G = (V, E, \Sigma, \lambda)$ 表示,V, E, Σ, λ 都如上文所述。如果一幅图 $G = (V, E, \Sigma, \lambda)$ 和另一幅图 $G' = (V', E', \Sigma', \lambda')$ 有 1-1 映射的关系 $f : V \to V'$，那么图 G 就是 G' 的*子图*，用 $G \in G'$ 表示。1-1 映射可以有这么几种

- 对于所有 $v \in V, \lambda(v) = \lambda'(f(v))$
- 对于所有 $(u, v) \in E, (f(u), f(v)) \in E'$
- 对于所有 $(u, v) \in E, \lambda(u, v) = \lambda'(f(u), f(v))$

换言之，如果一幅图保持着另一幅图的所节点标签，边关系，和边标签，那么这副图就是另一幅图的子图。一个*通路 (walk)* 是一个节点列表 $v_1, v_2, ..., v_n$, 对于所有 $i \in [1, n-1]$,v_i 和 v_{i+1} 有边连接。一个*路径 (path)* 是一个不含重复节点的通路, 即对于所有 $i \neq j, v_i \neq v_j$。

1.2 再生 Hilbert 空间

对于大量图数据的分析，核函数是一个很强大的处理工具。核函数的优势在于它无需精确计算对应点对就可以把一组数据放入一个高维的 Hibert 空间。我们用一个叫做*核*的函数来处理这个过程。一个二元函数 $K : X \times X \to \mathbb{R}$ 如果符合以下公式，则它是一个*半正定*函数。

$$\sum_{i,j=1}^{m} c_i c_j K(x_i, x_j) \geq 0 \tag{1}$$

上式中 $m \in \mathbb{N}$, 例子 $x_i \in X(i = [1, n])$, 系数集 $c_i \in \mathbb{R}(i = [1, n])$。另外，如果 $x, y \in \ X, K(x, y) = K(y, x)$ 那么这个二元函数就是**对称的**。一个对称半正定函数保证存在一个 Hilbert 空间 $\mathcal{H}$ 和一个映射 $\Phi : X \to \mathcal{H}$, 例如

$$k(x, x') = \langle \Phi(x), \Phi(x') \rangle \tag{2}$$

对于所有的 $x, x' \in X$。$\langle x, y \rangle$ 表示 $x\ y$ 的内积。这个结果就是我们所熟悉的 Mercer 定理。一个对称半正定函数又称为 Mercer 核函数简称为核函数。通过将数据空间变为 Hilbert 空间，核函数提供了一种对包括图在内的不同数据的统一分析环境。即使一开始的数据空间根本不像一个向量空间，也可以统一化。我们称这种归一化方法为" 核诀窍"，并将其应用在许多不同的数据分析任务上，包括分类，回归，通过准则分析的特征提取等。