Achieve Equal Influence on Attributes

——Elements of Statistical Learning, Chapter 14 Note

Shuai

December 29, 2013

The overall dissimilarities consists of dissimilarities from each attribute. To make the influence made by each attribute equal, one way is to assign weight to them.

We assume there are p attributes.

However, it is important to realise that setting the weight ω_j to the same value for each variable (say, $\omega = 1\ \forall\ j$) does *not* necessarily give all attributes equal influence. The influence of the jth attribute X_j on object dissimilarity $D(x_i, x_{i'})$ depends upon its relative contribution to the average object dissimilarity measure over all pairs of observations in the data set

$$\bar{D} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N} D(x_i, x_{i'}) = \sum_{j=1}^{p}\omega_j d_j$$

with

$$\bar{d_j} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N} d_j(x_{ij}, x_{i'j})$$

being the average dissimilarity on the jth attribute. Thus, the relative influence of the jth variable is $\omega_j \bar{d_j}$, and setting $\omega_j \sim 1/\bar{d_j}$ will give all attributes equal influence.

Take squared-error distance as an example, then

$$\begin{aligned}
\bar{d_j} &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}(x_{ij} - x_{i'j})^2 \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}(x_{ij}^2 - 2x_{ij}y_{i'j} + x_{i'j}^2) \\
&= \frac{\sum_{i=1}^{N} x_{ij}^2}{n} - 2E(X_{ij}X_{i'j}) + \frac{\sum_{i'=1}^{N} x_{i'j}^2}{n} \\
&= EX_j^2 - 2E(X_{ij}X_{i'j}) + EX_j^2 \\
&= 2EX_j^2 - 2E(X_j)^2 \\
&= 2Varj
\end{aligned}$$