

The following is a scheme for weighting variables in a mixed-type dataset to convert into a Gower matrix.

x_t is the t^{th} variable of dataset X .

If x_t is a numeric variable, then $w_t \leftarrow 1$.

Otherwise, let h_{tk} be the k^{th} column of the one-hot encoding H_t of x_t .

$$w_t \leftarrow \sum_k^{K_t} \widehat{\text{Var}}(h_{tk}) \left[1 - \sum_k^{K_t} \widehat{\text{Var}}(h_{tk}) \right] \widehat{\text{Kurt}}(H_t)$$

The first term is zero when all the values of the variable are the same, whereas the second term is zero when they're all different. The third term increases as the number of unique values increases.