Clustering Methods

Exercises 7/7, 22.5.2017

1. Several variants of SSW and SSB were tried for cluster validation, including SSW/k and k·SSW/SSB. Why not just k·SSW? Which of the following is most logical: k·TSE, k·MSE, k·nMSE.

2. Implement and test the measures mentioned above and test with A, S, DIM and Unbalance data. Report the results.

3. Knee-point detection was also considered using second derivative with the three sub-sequent clustering results as follows: $f(k$-$1) - 2f(k) + f(k$+$1)$, where f is any validity index such as TSE, BIC or the above WB-index. Any idea how exacHowever, it

4. Two different ways of using KNN links were presented for outlier detection: mean distance and the indegree. Can we combine these and how? Outline possible method.

5. Upload your answers to Sami. (samisi@cs.uef.fi) by 22.5. by 10.00 latest. In your email, have title "Clustering exercises 7/7".