Session 2: Applications of Cluster Analysis

# Question

❑ Can we use clustering to help with classification tasks?

  ❑ No! Clustering (unsupervised learning) and classification(supervised learning) are completely different paradigms and cannot be combined.

  ❑ Yes! We can use clustering as a preprocessing step in classification tasks to help us deal with feature sparsity issues as well as engineering complex features. For example, in a medical database, each patient may have a distinct real-valued measure for certain tests (e.g. glucose, cholesterol). Clustering patients first may help us understand how binning should be done on real-valued feature to reduce feature sparsity and improve accuracy on classification tasks such as survival prediction.

❑ Answer: Yes. As discussed in the lecture, clustering can be used as a preprocessing step in classification tasks.