

1 Kernel k-Means

Suppose we have a dataset $\{x_i\}_{i=1}^N, x_i \in \mathbb{R}^n$ that we want to split into K clusters. Furthermore, suppose we know a priori that this data is best clustered in a large feature space \mathbb{R}^m , and that we have a feature map $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^m$. How should we perform clustering in this space?

- (a) Write the objective for K-means clustering in the feature space (using the squared L_2 norm in the feature space). Do so by explicitly constructing cluster centers $\{\mu_k\}_{k=1}^K$ with all $\mu_k \in \mathbb{R}^m$.
- (b) Write an algorithm that minimizes the objective in (a).
- (c) Write an algorithm that minimizes the objective in (a) without explicitly constructing the cluster centers $\{\mu_k\}$. Assume you are given a kernel function $\kappa(x, y) = \phi(x) \cdot \phi(y)$.