

Introduction to Machine Learning

Lecture 4: Clustering

Alexis Zubiolo

`alexis.zubiolo@gmail.com`

Data Science Team Lead @ Adcash

November 15, 2016

k -means

We look for a partition $S = \{S_1, S_2, \dots, S_k\}$ minimizing the within-cluster sum of squares.

$$\arg \min_S \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|_2^2$$

where

$$\mu_i = \frac{1}{|S_i|} \sum_{x \in S_i} x$$

is the mean of points in S_i