

ML Section 3: Unsupervised Learning

Adam Hawley

March 14, 2019

Contents

1	Introduction	1
2	Clustering	1
2.1	k-means Algorithm	1
3	Summary of ML	2

1 Introduction

- Supervised learning requires pre-categorised training data provided by a *supervisor*.
- Reinforcement learning requires a reward function which is normally provided by an engineer.

What if only data without any annotation is available? This can happen in certain circumstances such as behavioural segmentation, e.g. segmentation of customers by purchase history, inventory categorisation and detecting anomalous behaviour.

2 Clustering

2.1 k-means Algorithm

Input: data (in vector format) and a number of clusters k .

1. Randomly select k centroids c_i from the data.
2. Assign each data point x to a centroid according to: $(\operatorname{argmin}_{c_i \in C}) \operatorname{dist}(c_i, x)^2$

3. Update centroids, by setting the centroid of each cluster S_i to the mean of all data points assigned to the cluster: $c_i = \frac{1}{||S_i||} \sum_{x_i \in S_i} x_i$.
4. Repeat steps 2 and 3 until convergence (or had enough).

How to chose k ? When you plot the “Average within-cluster distance to centroid” vs. k there will be an *elbow point* where increasing k no longer decreases the distance by much. This is the ideal value of k .

3 Summary of ML

- Supervised learning:
 - Decision tree learning
 - Neural networks (backpropagation)
- Reinforcement learning:
 - Q Learning
- Unsupervised learning:
 - K-Means Clustering