Unsupervised Learning

Notes by José A. Espiño P. [1]

Summer Semester 2022–2023

[1] The content in these notes is sourced from what was covered in the MOOG the document is named after. I claim no autorship over any of the contents herein.

Contents

1 Unsupervised Learning

As opposed to supervised learning, where the training set is a combination of inputs and target labels, unsupervised learning consists in just inputting unlabeled data and asking the algorithm to find interesting patterns in it.

1.1 Clustering

A clustering algorithm looks at a number of data points and automatically finds which data points are related to one another, by trying to find clusters.
The K–means Algorithm is the most commonly used clustering algorithm. It consists of repeatedly the following things:

1. Take a random guess at where are the centres of each cluster (the cluster centroids)
2. Go through each other sample and assign it to its closest centroid
3. Look at all the points in each cluster, calculate the average of them, and assign a new centroid

There will stop when it reaches convergence, namely, when there are no changes in the location of the cluster centroids or when there is no change to the assignment of every other datapoint.
A formal definition of the k–means algorithm is as follows:

1. Randomly initialise k cluster centroids $\mu_1, \mu_2, \ldots, \mu_k$ Each μ_n is a vector with the same dimensions as the training examples.
2. for $i = 1$ to m, where m is the amount of training samples, we set $c^{(i)}$:= index (from 1 to k) of the cluster centroid closest to $x^{(i)}$. Mathematically, this is obtained b y computing the distance between points, $||||x^{(i)} - \mu_n||||^2$. You want to find the value of n that would minimise such distance. This value of n is what would be assigned to $c^{(i)}$.
3. Update the location of the cluster centroids to be the average or the mean of the points assigned to the cluster n. Mathematically, this involves computing the average on every dimension, which will give you the new mean. If the cluster is empty, the algorithm just eliminates the cluster.

The k–means algorithm is actually trying to optimise a certain cost function:

$$J(c^{(1)}, ..., c^{(m)}; \mu_1, ..., \mu_k) = \frac{1}{m} \sum_{i=1}^{m} ||||x^{(i)} - \mu_{c^{(i)}}||||^2$$

1.2 Anomaly Detection

2 Recommender Systems

3 Reinforcement Learning

```
sample!
```