K-MEANS, EM ALGORITHM AND VARIATIONAL METHODS

MAHESH VEMULA[1]

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

We will closely study three well known clustering methods and understand the inter -relations between these methods.

1. K-Means
2. E-M (Expectation Maximization)
3. Variational Inference

We will then later explore these methods on some corpora based datasets.

2 TERMINOLOGY

Given a set of m samples $x_1, x_2, ...x_m$, we wish to

1. Find K clusters such that within cluster sum of squares distances is minimized.

$$J = \sum_{k=1}^{K} \sum_{x_i \in C_k} \|\mathbf{x}_i - \boldsymbol{\mu}_k\|^2 \tag{1}$$

2. Fit the parameters of a probabilistic model with $p(\mathbf{x}, \mathbf{z})$ being the joint likelihood.

K-Means algorithm solves the first step, while variational methods solve the second part. The EM algo is a special case of variational methods. We shall investigate these two algos in greater detail now.

3 EM ALGORITHM

The standard example when introducing EM algorithm is the Gaussian Mixture problem. The assumed model is that the samples are drawn from a mixture Gaussian distribution whose parameters are unknown.

Consider a Gaussian Mixture model of the following form

$$p_1\mathcal{N}(\boldsymbol{\mu}_1, \Sigma_1) + p_2\mathcal{N}(\boldsymbol{\mu}_2, \Sigma_2) + \cdots + p_K\mathcal{N}(\boldsymbol{\mu}_K, \Sigma_K) \tag{2}$$

Let us consider the following generative model for the sample dataset

1. Draw a random variable $z_i = k$ from $1 \ldots k$. This variable indicates from which distribution the sample will be drawn.
2. Draw a sample $\mathbf{x}_i$ from the k^{th} distribution $\mathcal{N}(\boldsymbol{\mu}_k, \Sigma_k)$

In our tasks we have to infer the parameters of the distribution (2) from the samples $x_1, x_2, ...x_m$. I will follow Prof Andrew Ng's exposition in deriving the EM algorithm steps.

3.1 *Jensen Inequality*

The Jensen Inequality states that if f is a **convex** function then we have

$$E[f(\mathbf{X})] \geqslant f[E(\mathbf{X})] \tag{3}$$

We have the converse result for a **concave** function. log is a concave function therefore we have

$$E[\log(\mathbf{X})] \leqslant \log[E(\mathbf{X})] \tag{4}$$

3.2 *EM Steps*

The EM is a general recipe for solving maximum likelihood problems in the presence of hidden variables. In the context of estimating mixtures, the hidden variables z_i are the random variables that indicate from which distribution the sample was drawn.

We begin by expressing the log-likelihood of the data as follows. The parameters of estimation include the $\Theta = \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i$.

$$\begin{aligned} l(\theta) &= \sum_{i=1}^{m} \log p(x_i; \Theta) && (5) \\ &= \sum_{i=1}^{m} \sum_{z_i=1}^{k} \log p(x_i, z_i; \Theta) && (6) \end{aligned}$$

We introduced the following important trick in the form of a proposal distribution $Q(z_i)$ as follows

$$\begin{aligned} l(\theta) &= \sum_{i=1}^{m} \sum_{z_i=1}^{k} \log p(x_i, z_i; \Theta) && (7) \\ &= \sum_{i=1}^{m} \sum_{z_i=1}^{k} \log \frac{Q(z_i) * p(x_i, z_i; \Theta)}{Q(z_i)} && (8) \\ &= \sum_{i=1}^{m} \sum_{z_i=1}^{k} \log Q(z_i) \frac{p(x_i, z_i; \Theta)}{Q(z_i)} && (9) \\ & && (10) \end{aligned}$$