

5.3 von Mises Fisher model

08t4023y 金田昌侑

von Mises-Fisher 分布とは

- 直接的なデータとしてのガウス分布の類似度である。
- 4章のspherical kmeansのアルゴリズムはv MF分布に従う。
- 他のモデルよりもテキストデータの分布に近づく。

vMF分布の式

$$p(d | \theta) = \frac{1}{Z(\Sigma)} \exp\left(\sum \frac{d^T \mu}{|\mu|}\right)$$

d はコーパス D 内のコーパス、 θ は終了状態の内のコーパス。 Z はベッセル関数を意味する。

vMF分布の特徴

元となる分布として、このモデルを使っているemアルゴリズムは、 Σ がM-stepで計算されている。

このモデルを使っているemアルゴリズムはspherical k-meansよりも性能が優れている。

5.4 Model-based k-means

Model-based k-meansとは

- emアルゴリズムに基づいたk-means
- 評価と割り当てを繰り返し、その回数は $O(kn)$ となる。

Model-based k-meansの特徴

- このモデルと他のモデルとの違いは、クラスターごとにそれぞれの要素を与えず、後の確立に基づいたモデルを教えないことである。