

EM algorithm

Racha Amina DJAGHLOUL

January 31, 2024

1 Principe de l'Algorithme EM :

L'algorithme EM est une méthode itérative utilisée pour estimer les paramètres d'un modèle statistique en présence de données incomplètes ou de variables latentes.

2 Les Étapes de l'Algorithme EM :

1. Étape d'Estimation (E-step) :

Concept : On essaie de deviner à quel groupe chaque observation pourrait appartenir, en fonction de ce qu'on sait déjà.

Calcul : On fait des calculs des probabilités d'appartenance pour mesurer à quel point chaque observation a de chances d'être dans chaque groupe. Cela se fait en considérant les centres des classes, les probabilités de classe, et les matrices de variance-covariance des clusters.

Rôle : On assigne des "poids" à chaque observation pour montrer à quel point elle pourrait être liée à chaque groupe. En gros, on fait des estimations sur la probabilité que chaque observation soit dans chaque groupe.

2. Étape de Maximisation (M-step) :

Concept : Cette étape vise à ajuster les paramètres du modèle pour maximiser la probabilité des données observées, compte tenu des probabilités d'appartenance aux clusters estimées à l'étape précédente.

Mise à Jour des Paramètres : On met à jour les probabilités de classe, les centres des classes, et les matrices de variance-covariance des clusters en fonction des probabilités estimées à l'étape précédente.

Rôle : Cette étape "améliore" nos estimations initiales en ajustant les paramètres du modèle en fonction des informations récemment acquises sur les probabilités d'appartenance aux clusters. C'est un peu comme peaufiner nos choix pour qu'ils collent mieux à la réalité des données qu'on a observées.

3 Application des Étapes dans le Script :

- Initialisation des Paramètres (Étape Hors-Code) : Avant le script, on initialise les paramètres tels que les centres des classes, les probabilités de classe, et les matrices de variance-covariance des clusters.
- Étape d'Espérance (E-step) dans le Script : Rôle : Estimation des probabilités d'appartenance aux clusters pour chaque observation.
- Étape de Maximisation (M-step) dans le Script : Rôle : Mise à jour des probabilités de classe, des centres des classes, et des matrices de variance-covariance des clusters.
- Convergence (Étape Hors-Code) : Avant le script : L'algorithme itère ces étapes jusqu'à ce que les changements deviennent négligeables, indiquant la convergence.