2024（令和６）年度 九州大学大学院経済学府修士課程第Ⅰ期入学試験問題
（一般選抜）

# 情　報　管　理

次の２問から，１問を選択し解答しなさい。

問１

クラスタリングは，$d$ 個の特徴量を持つ $n$ 個のデータ $x_1, x_2, \cdots, x_n \in \mathrm{R}^d$ を互いに素な $K$ 個の集合（クラスタ）に分割するタスクである。最も代表的なクラスタリング手法の一つである $K$-means は，あらかじめクラスタの個数 $K$ を定め，$n$ 個のデータをランダムにクラスタ 1 から $K$ のいずれかに割り当て，以下の 2 ステップ

1. クラスタ $k = 1, \cdots, K$ に対して，それぞれのクラスタに含まれるデータの中心 $\mu_k$ を求める
2. $n$ 個のデータに対して，最も近い中心 $\mu_k$ に対応するクラスタ $k$ にデータ $x_i$ を割り当てる

を繰り返し，クラスタの割当に変化がなくなった時点で終了するアルゴリズムである。

各データのクラスタへの割当が固定された下で，クラスタ $k$ のスコア $S_k$ を，中心 $\mu_k$ からそのクラスタに含まれる各データ $x_i$ との（ユークリッド）距離の二乗和

$$S_k = \sum_{i \in C_k} \|x_i - \mu_k\|^2$$

で定義したときに，$S_k$ が最小となる中心 $\mu_k$ は，そのクラスタに含まれるデータの平均ベクトル，つまり

$$\mu_k = \frac{1}{|C_k|} \sum_{i \in C_k} x_i$$

となることを説明しなさい。ただし，$C_k$ はクラスタ $k$ に含まれるデータの集合とし，$|C_k|$ はその要素数とする。

問２

情報推薦システムは，膨大な情報の中からユーザの嗜好にあった情報を提供するシステムで，その実現方式は，内容に基づくフィルタリングと協調フィルタリングに大別される。協調フィルタリングは，他のユーザの嗜好のデータ（アイテムに対する評価）を使い，対象となるユーザの各アイテムに対する評価を予測し，予測値が高いアイテムを推薦する。ユーザベースの協調フィルタリングでは，どのようにしてアイテムに対する評価を予測しているのか説明しなさい。