

情報管理

次の2問から、1問を選択し解答しなさい。

問1  $n$  を正の整数、 $S \subseteq \mathbb{R}^n$  を凸集合とし、 $f: S \rightarrow \mathbb{R}$  を凸関数とする。このとき、以下の問題に答えよ。

- (1)  $S$  が凸集合であることの定義を述べよ。
- (2)  $f: S \rightarrow \mathbb{R}$  が凸関数であることの定義を述べよ。
- (3)  $x^*$  が  $f$  の局所的最小解であることの定義を述べよ。
- (4)  $x^*$  を  $f$  の局所的最小解とする。このとき、 $x^*$  は  $f$  の大域的最小解になることを証明せよ。

問2 主成分分析とは、多くの特徴量を持つ複雑なデータを、より少ない指標で要約するための次元圧縮手法の1つである。いま、 $m$  次元の特徴を持つ  $n$  個のデータ  $\{x \in \mathbb{R}^m \mid i = 1, 2, \dots, n\}$  が与えられたときに、主成分分析における第1主成分および第2主成分を求める最適化問題をそれぞれ定式化しなさい。その上で、それぞれの最適化問題に課される制約条件の必要性について述べ、最終的にどのように主成分が与えられるか説明しなさい。以降では、 $n$  個のデータを各行に並べた行列を

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \dots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix} \in \mathbb{R}^{n \times m}$$

で表し、行列  $X$  は中心化されている、つまり各特徴量の標本平均は0であると仮定する。ただし、 $m$  と  $n$  は共に2以上の正の整数とする。