

問題発見技法 再試験課題

【課題】

要素数がそれぞれ 5 個ある①～⑩の 10 個の対象についてのデータ・ファイル "2006reexam.csv" がある。この 10 個の対象についてクラスター分析を行う。距離はユークリッド平方距離 (squared Euclidean distance) を用い、ウォード法 (Ward method) でクラスター化を行う。授業時の PPT ファイル (P.33～36, 表 7-2～7-7) を参考に、最初の距離を求めた表、及び、クラスター化の計算過程全てを示した表を Excel 等で作成すること。なお、作成した表毎に一つずつ具体的な計算式を示せ。

計算式の示し方の例：表 7-2 から表 7-3 を作る際の一つの計算式。②と⑤をクラスター化して「②2」とし、「②2」と①との距離を重心法で測る場合

$$s_{\textcircled{2}2,\textcircled{1}} = \frac{1}{1+1}13 + \frac{1}{1+1}10 - \frac{1 \cdot 1}{(1+1)^2}1 = \frac{45}{4} (=11.25)$$

最後に樹形図 (デンドログラム) を描く (Word, Excel の「オートシェイプ」や Illustrator, その他お絵かきソフトを利用すること)。また、この樹形図から、10 個の対象を 3 つのクラスターに分類する場合はどこに線を引いたらいいかも、同じ図の中で示すこと。

提出するレポートは以下の要領で、「1. 分析対象のデータ」～「4. 意見・コメント」の順に書くこと。なお、表紙として 1 ページ「タイトル」を付けること。科
目名、提出日、氏名、学籍番号など書くべきことをこの表紙に書く。

1. 分析対象のデータ ← 与えられたデータの表を書く (授業時 PPT ファイル P.33 表 7-1 を参考にせよ)
2. クラスター化 ← 上記に示したクラスター分析を行い、その過程・結果をここに掲載する (レポートにするのだから、上記表や式だけでなく、前後の文章等必要な肉付けは自分で考えて書くこと。単に表や式だけが書いてあって、それが何を意味するのか何も説明がない場合は不可)
3. 樹形図 ← クラスター分析の結果得られた樹形図を描き、3 つにクラスター化した場合どうなるかを示す。樹形図は 1 ページ以内に納めること。
4. 意見・コメント ← クラスター分析に対する意見・コメントを書く。何を書いても良いが、単なる感想は不可。

【提出について】

提出期限：9 月 16 日 (土) 21 : 00 【時間厳守】

提出宛先：メールにて khotta@shonan.bunkyo.ac.jp

提出物：作成したレポートを PDF ファイルにして、メールに添付し提出すること。PDF ファイル以外は受け付けないので注意。