

先端データ解析論 第 10 回小レポート

情報理工学系研究科電子情報学専攻 M1 堀 紡希 48216444

2021 年 6 月 17 日

宿題 1

別の ipynb ファイルで提出します。

宿題 2

クラス内

$$\begin{aligned} S^{(w)} &= \sum_{y=1}^c \sum_{i: y_i=y} (\mathbf{x}_i - \boldsymbol{\mu}_y)(\mathbf{x}_i - \boldsymbol{\mu}_y)^\top \\ &= \sum_{y=1}^c \sum_{i: y_i=y} (\mathbf{x}_i - \frac{1}{n_y} \sum_{i': y_{i'}=y} \mathbf{x}_{i'}) (\mathbf{x}_i - \frac{1}{n_y} \sum_{i': y_{i'}=y} \mathbf{x}_{i'})^\top \\ &= \frac{1}{2} \sum_{y=1}^c \sum_{i, i': y_i=y_{i'}=y} (\frac{1}{n_y} \mathbf{x}_i - \frac{1}{n_y} \mathbf{x}_{i'}) (\frac{1}{n_y} \mathbf{x}_i - \frac{1}{n_y} \mathbf{x}_{i'})^\top \\ &= \frac{1}{2} \sum_{i, i'=1}^n Q_{i, i'}^{(w)} (\mathbf{x}_i - \mathbf{x}_{i'}) (\mathbf{x}_i - \mathbf{x}_{i'})^\top \end{aligned}$$

より示された. 二行目から三行目で各要素についての和をそれぞれの index 順に取り直している. i と i' の組み合わせは二度計算されるので $1/2$ 倍する必要がある.

クラス間

散布行列は, 中心化の仮定から得られる

$$\sum_{i=1}^n \mathbf{x}_i \sum_{i'=1}^n \mathbf{x}_{i'}^\top = \mathbf{0}$$

を用いると以下のように表せる.

$$\begin{aligned} \boldsymbol{C} &= \sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{x}_{i'}^\top \\ &= \frac{1}{2} \sum_{i,i'=1}^n \frac{1}{n} (\boldsymbol{x}_i \boldsymbol{x}_i^\top + \boldsymbol{x}_i \boldsymbol{x}_{i'}^\top) \\ &= \frac{1}{2} \sum_{i,i'=1}^n \frac{1}{n} (\boldsymbol{x}_i - \boldsymbol{x}_{i'}) (\boldsymbol{x}_i - \boldsymbol{x}_{i'})^\top \end{aligned}$$

クラス内の結果と, $\boldsymbol{S}^{(b)} = \boldsymbol{C} - \boldsymbol{S}^{(w)}$ を用いると,

$$\boldsymbol{S}^{(b)} = \frac{1}{2} \sum_{i,i'=1}^n Q_{i,i'}^{(b)} (\boldsymbol{x}_i - \boldsymbol{x}_{i'}) (\boldsymbol{x}_i - \boldsymbol{x}_{i'})^\top$$

となる.