

第 10 章

モデル推定

10.1 目的

教師なし学習であるモデル推定について学びます。Xmean クラスタリングを行うためには、Weka に外部パッケージをインストールする必要があります。インストール手順は教科書の付録 A を参照してください。

10.2 Weka HierarchicalClusterer, SimpleKMeans, XMeans, EM

実践演習 10-1

例題 10.1, 10.2, 10.3, 10.5 に従い、iris.2D データのクラスタリング（階層的・kMeans・XMeans・EM）を行い、結果を確認せよ。

10.3 Weka による異常検出

実践演習 10-2

例題 10.4 に従い、iris データの一部を異常値に書き換え、LOF フィルタで検出を行え。その際、異常値とする値をいくつか試し、どの程度なら異常と判定できるか検討せよ。

10.4 sklearn cluster

sklearn では、クラスタ数を自動的に決定してくれるアルゴリズムとして Affinity Propagation（クラスタ中心になるべきパラメータ (responsibility) とクラスタへの属しやすさ (availability) をデータ間で伝達して収束させる手法）が用意されています。

実践演習 10-3

scikit-learn 付属の iris データから petallength と petalwidth 特徴を抜き出して 2 次元教師なしデータを作成し、階層的クラスタリング AgglomerativeClustering、k-means クラスタリング KMeans、クラスタ数決定の自動化 Affinity Propagation、確率密度推定 GaussianMixture を用いてクラスタリングを行え。