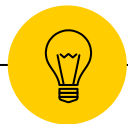


Hiddenによるデータ分析

- 葉について -





データについて

UCI Machine Learning Repository

異なる植物種に由来する、葉のデジタル画像から抽出された形状やテクスチャ集合データ

<https://archive.ics.uci.edu/ml/datasets/Leaf>



データについて

調査基準

- ❑ 葉についての情報12成分を抜粋し軸にした。
- ❑ 340の様々な葉のデータを折れ線にした。



データについて

調査品目一覧

※調査品目名は独断による。

離心率 (Eccentricity)	アスペクト比 ⁽¹⁾ (Aspect Ratio)	離角 (Elongation)	占有率 (Solidity)
湾曲係数 (Isoperimetric Factor)	最大凹み深さ (Maximal Indentation Depth)	突出さ (Lobedness)	平均強度 (Average Intensity)
標準偏差 (Average Contrast)	滑らかさ (Smoothness)	均一性 (Uniformity)	エントロピー (Entropy)

(1) 長辺:短辺



データについて

補足

→ 離角(Elongation)

$1 - 2d_{\max}/D(l)$ で計算される。

ただし $2d_{\max}/D(l)$ は最大の内接円の直径と最小の外接円の直径との間の割合。

つまり、外接円から内接円を除いた部分の割合を表す。



データについて

補足

→ 占有率(Solidity)

$A(I)/A(H(I))$ で計算される。

ただし $A(I)$ は葉 I の面積、 $A(H(I))$ は葉 I の凸包。

つまり、葉が凸形状にどの程度収まっているかを表す。



データについて

補足

→ 湾曲係数(Isoperimetric Factor)

$4\pi A(I)/L(\partial I)^2$ で計算される。ただし $L(\partial I)$ は葉 I の面に対する輪郭の長さ。曲がりくねった輪郭の葉ほど値が小さくなる。



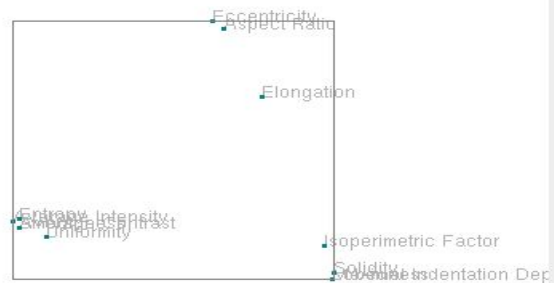
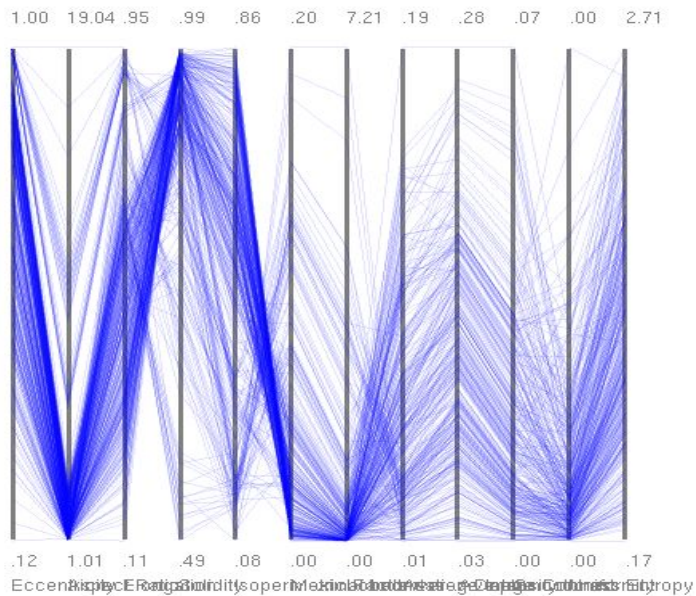
データについて

補足

- 標準偏差(Average Contrast)
葉の強度の標準偏差(散らばり具合)を表す。
- エントロピー(Entropy)
強度のランダム性の尺度を表す。



実行結果

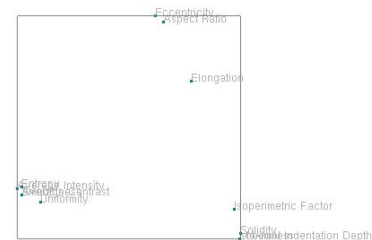
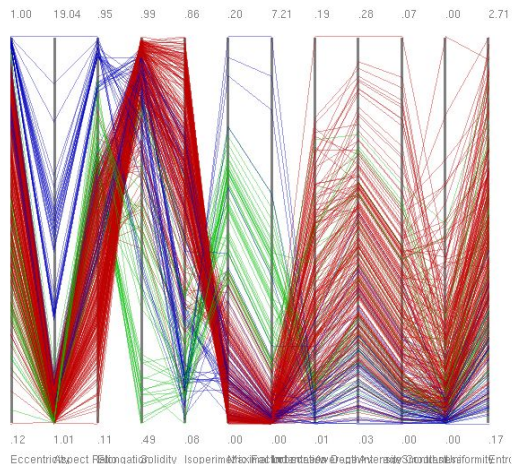




クラスタリング

Clustering = 3のときにある程度分類できた。

散布図から、成分同士の関連性が見られる。

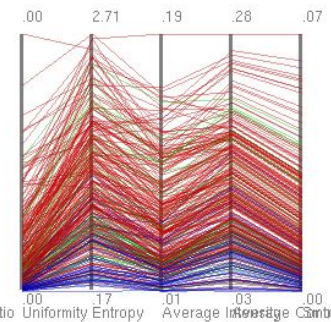
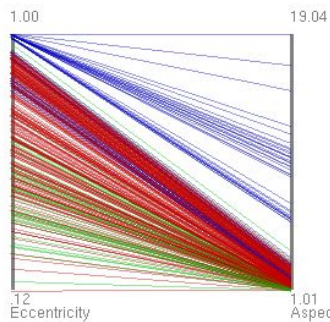
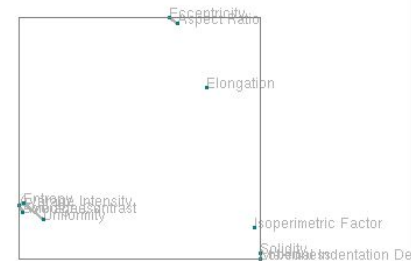
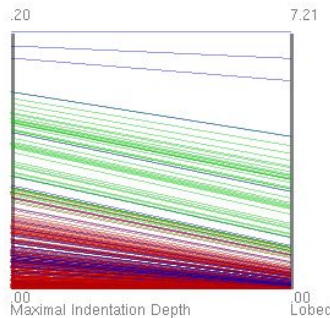




相関観察①

正の相関

1. 最大凹み深さと突出さ
2. 離心率とアスペクト比
3. 均一性とエントロピーと平均強度と標準偏差





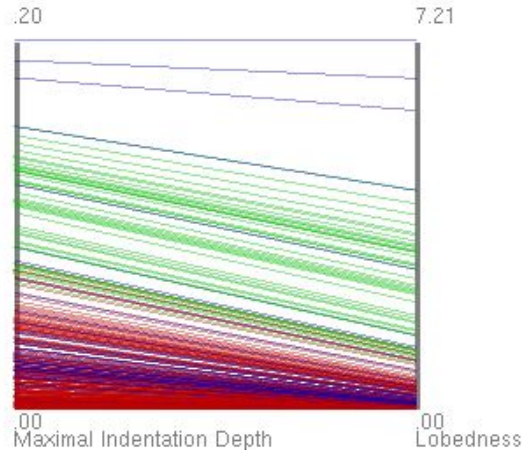
相関観察①

正の相関

1. 最大凹み深さと突出さ

→ きれいな正の相関が見られた。

凹みが大きい分突出しているということなので正の相関が成り立つ。



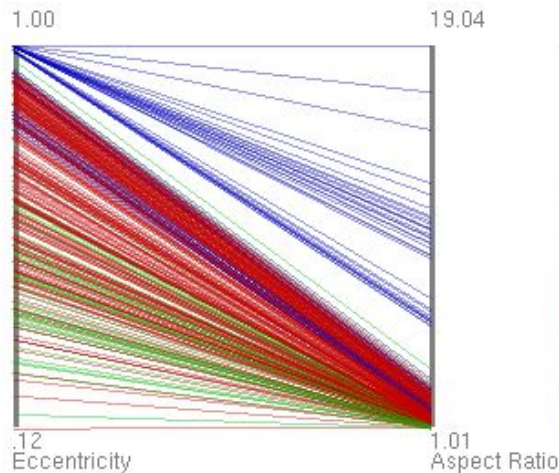


相関観察①

正の相関

2. 離心率とアスペクト比

→ 縦横比が大きいほど中心から離れているということなので、正の相関が成り立つ。

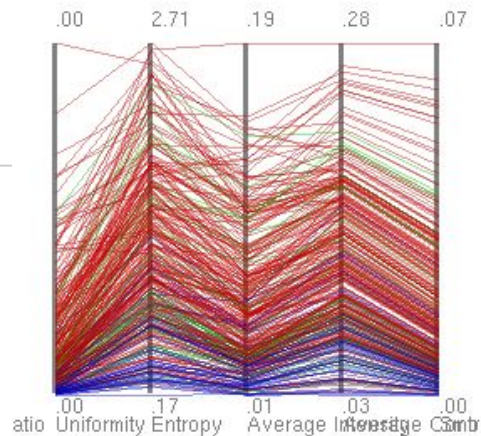




相関観察①

正の相関

3. 均一性, エントロピー, 平均強度, 標準偏差, 滑らかさ
- 全て強度に対する成分情報であるため、いずれか一方の成分が増加すると他方の成分も増加するように、正の相関が成り立つ。





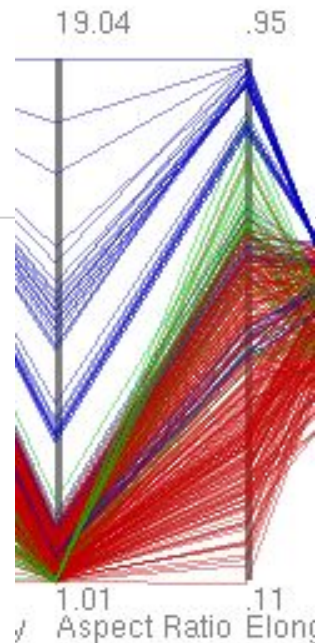
相関観察②

正の相関

アスペクト比と離角

→ 右下の8のように縦横比が大きいほど
外接円から内接円を引いた部分の割合
が大きいので正の相関。

4はその逆を表す。



8



4



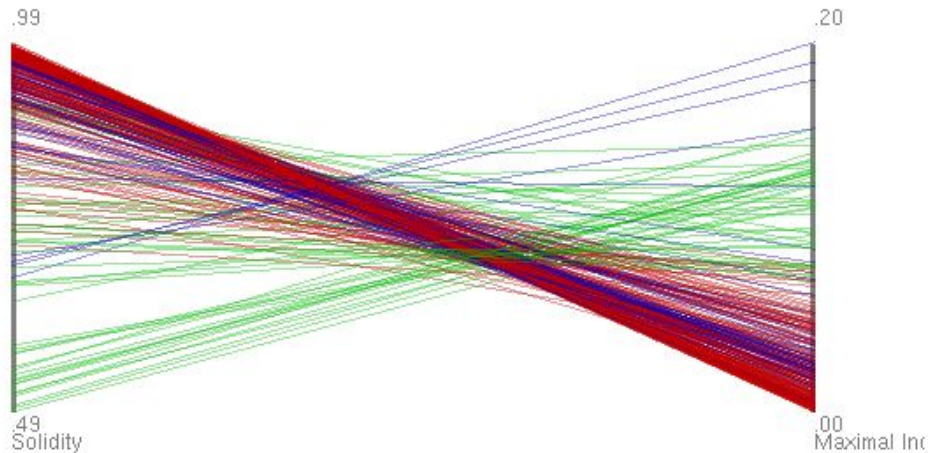


相関観察③

負の相関

占有率と最大凹み深さ

- 右下の38のように、輪郭がギザギザして占有率が小さいほど凹み深さの最大値は大きくなる。
27はその逆を表す。



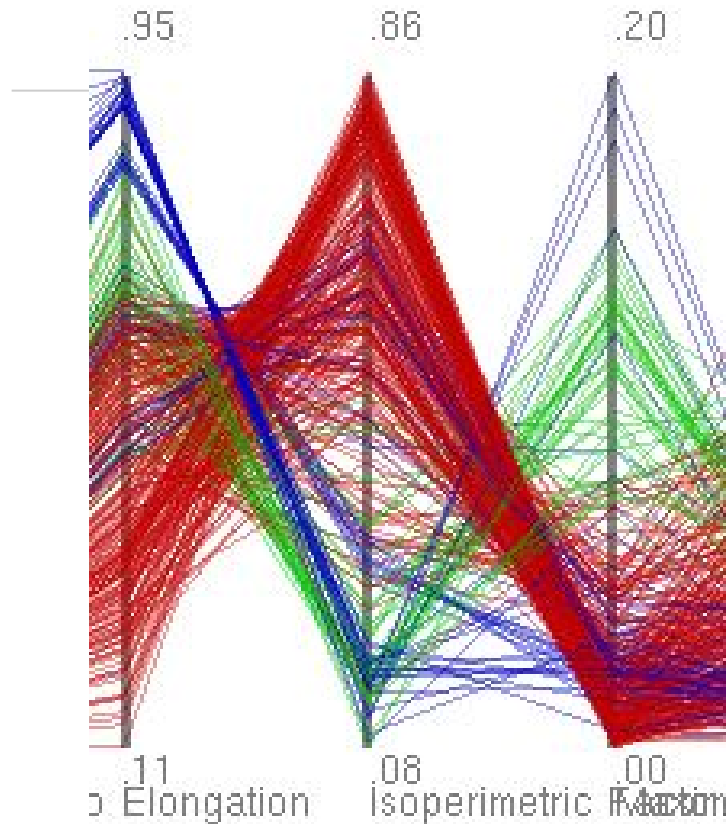


相関観察④

負の相関

離角と湾曲係数と最大凹み深さ

- 曲がりくねった葉ほどもともと絡まっているので凹みも深いので負の相関になる。
- また、離角が小さいほど曲がったり絡まったりしにくいことがわかる。



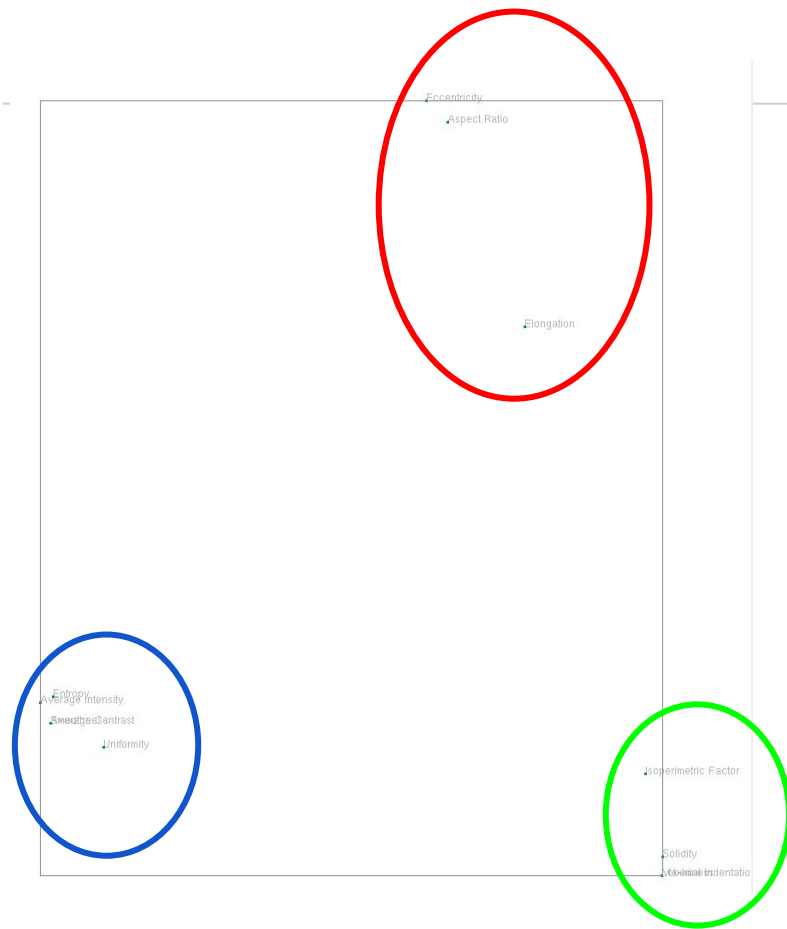


相関観察⑤

散布図

大きく分けて3つに見える。

最大凹み深さと突出さ、標準偏差と滑らかさはそれぞれ散布図上では完全に一致していた。

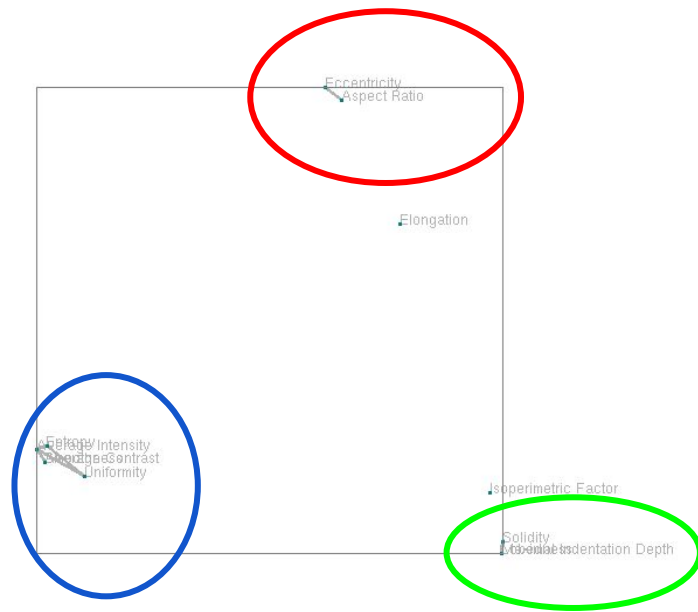




相関観察⑥

散布図

Hiddenにて右上のバーを動かすと
まず3箇所に線が引かれた。



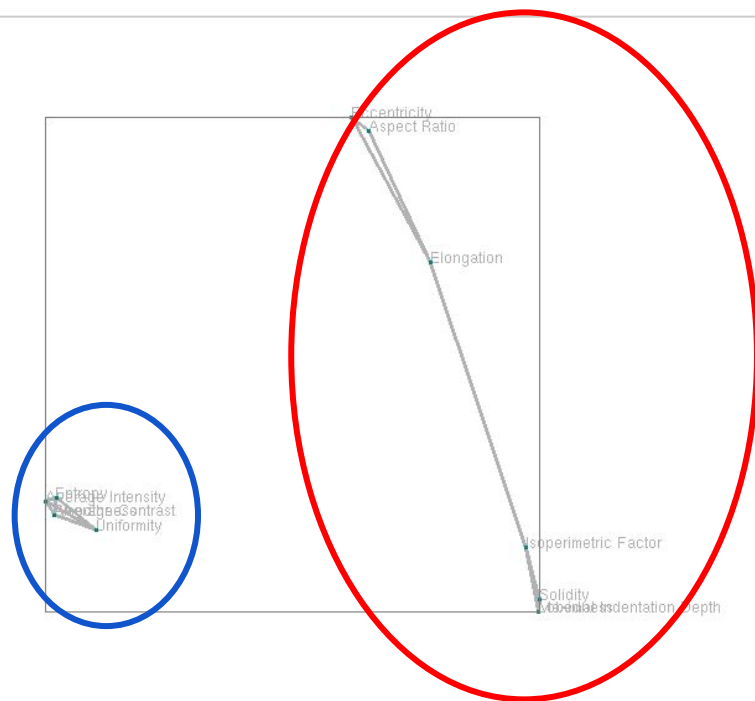


相関観察⑦

散布図

さらに右上のバーを動かすとさらに2分割された。

青枠の部分はこれまでの考察結果から特に正の相関が強い。





まとめ

- ❖ 大体3つにクラスタリングできているように見えた。
- ❖ たくさんある植物の葉にも正負の相関がはっきりと見られた。
- ❖ 新たな発見としては、離角が小さいほど葉は曲がったり絡まったりしにくいことがわかった。



感想

軸となるデータの属性について値の計算方法を理解するのに苦労しましたが、相関に対する考察がしやすく、楽しんで取り組みました。

上京してから自然と関わることが少なくなったため、久しぶりにたくさんの葉のデータを見られてリラックスできました。



参考文献

'Evaluation of Features for Leaf Discrimination',
Pedro F.B. Silva, Andre R.S. Marcal, Rubim M.
Almeida da Silva (2013). Springer Lecture Notes in
Computer Science, Vol. 7950, 197-204.