Unsupervised Classification

Allan Aasbjerg Nielsen
alan@dtu.dk

Technical University of Denmark
DTU Compute – Applied Mathematics and Computer Science

27 Feb 2019

DTU

DTU

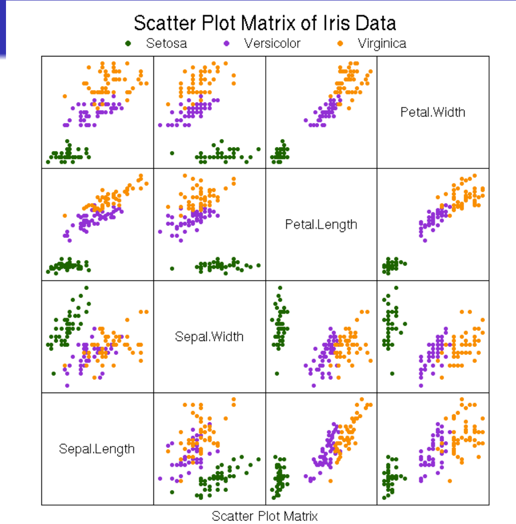
Classification

- classification is the process of grouping observations (pixels or regions) into classes intended to represent different physical objects or types
- here, the production of a **thematic map** from (image) data with digital numbers representing for example reflected or emitted EM-radiation in different wavelength bands
- very many classification methods ranging from quite simple to highly advanced
- two major groups of methods: supervised and unsupervised
  - supervised: ideally physical classes but not necessarily statistically distinct
  - unsupervised: statistically distinct but not necessarily physical classes

DTU

Feature space

1. p variables
   C classes
   N observations (or samples)
2. $x_i,\ i = 1, \ldots, N,\ p \times 1$
   is a point (or vector) in
   p-dimensional **feature space**
3. figure shows all possible pairwise projections on original variables



DTU

K-means

1. choose C
2. assign C class centres $\boldsymbol{\mu}_c$
3. calculate distance, e.g., $D^2_{Eic} = (\boldsymbol{x}_i - \boldsymbol{\mu}_c)^T(\boldsymbol{x}_i - \boldsymbol{\mu}_c)$ for all observations to all class centres, $i = 1, \dots, N,\ c = 1, \dots, C$
4. assign class c to $\boldsymbol{x}_i$ if distance smallest for class c
5. compute new class centres $\boldsymbol{\mu}_c$ (include only obs in class c)
6. iterate from third step

Initialization of $\boldsymbol{\mu}_c$

1. random observations within range of data
2. first C 'different enough' observations
3. based on PCA, e.g., uniformly distributed along first PC axis, or in plane spanned by two first PC axes
4. ...

Fuzzy c-means

1. choose C
2. assign C class centres $\boldsymbol{\mu}_c$
3. calculate distance, e.g., $D^2_{Eic} = (\boldsymbol{x}_i - \boldsymbol{\mu}_c)^T(\boldsymbol{x}_i - \boldsymbol{\mu}_c)$ for all observations to all class centres
4. assign degree of membership u_{ic} to $\boldsymbol{x}_i$ for all classes, e.g., $u_{ic} = (1/D^2_{Eic})/\sum_{j=1}^C 1/D^2_{Eij}$ leading to $\sum_{c=1}^C u_{ic} = 1$
5. compute new class centres (include all obs weighted by u_{ic}) $\boldsymbol{\mu}_c = \sum_{i=1}^N u_{ic}\boldsymbol{x}_i / \sum_{i=1}^N u_{ic}$
6. iterate from third step

Unsupervised: Gaussian mixture models, GMM

1. Bayes' rule: $P(\omega_c|\boldsymbol{x}_i) = K\ P(\boldsymbol{x}_i|\omega_c)P(\omega_c)$ with $1/K = \sum_{j=1}^C P(\boldsymbol{x}_i|\omega_j)P(\omega_j)$
2. **GMM**: Given some $u_{ic} = P(\omega_c|\boldsymbol{x}_i)$ with $\sum_{c=1}^C u_{ic} = 1$, calculate
3. $P(\omega_c) = \frac{1}{N}\sum_{i=1}^N u_{ic}$ (here the mixing proportion of class c)
   $\boldsymbol{\mu}_c = \frac{1}{NP(\omega_c)}\sum_{i=1}^N u_{ic}\boldsymbol{x}_i$
   $\boldsymbol{\Sigma}_c = \frac{1}{NP(\omega_c)}\sum_{i=1}^N u_{ic}(\boldsymbol{x}_i - \boldsymbol{\mu}_c)(\boldsymbol{x}_i - \boldsymbol{\mu}_c)^T$
4. $\boldsymbol{\mu}_c$ and $\boldsymbol{\Sigma}_c$ define $P(\boldsymbol{x}_i|\omega_c)$ which with $P(\omega_c)$ via Bayes' rule give a new $u_{ic} = P(\omega_c|\boldsymbol{x}_i)$ which in turn gives a new $P(\omega_c)$: iterate
5. example on Expectation Maximization (EM) algorithm
   E-step: calculate $P(\omega_c)$, $\boldsymbol{\mu}_c$, $\boldsymbol{\Sigma}_c$
   M-step: calculate $P(\omega_c|\boldsymbol{x}_i)$ in Bayes' rule

Initialization of μ_c and Σ_c

1. select observations at random as initial means
   mixing proportions are uniform
   initial covariance matrices are diagonal, elements on the diagonal are the variances
2. start with result from k-means or fuzzy c-means
3. ...

Hierarchical clustering

1. hierarchical clustering groups data over a variety of scales by creating a cluster tree or dendrogram
2. the tree is not a single set of clusters, but rather a multilevel hierarchy, where clusters at one level are joined as clusters at the next level
3. this allows you to decide the level or scale of clustering that is most appropriate for your application
4. two extremes: every pixels is its own cluster vs entire image is one cluster

Matlab

1. Statistics and Machine Learning Toolbox
2. Cluster Analysis
3. k-means: `kmeans`
4. GMM: `fitgmdist`, `cluster`, `posterior`

Matlab exercise

1. Experiment with Matlab's implementations of k-means (`kmeans`) and Gaussian Mixture Models, GMM (`fitgmdist, posterior`). Try different numbers of clusters, different initializations (option '`Replicates`'), use original variables and first few principal components, etc. Apply to igalmss data.
2. Write a small report or a readable journal (3-4 pages) including figures and Matlab code.