02450: Introduction to Machine Learning and Data Mining

Mixture models and density estimation

DTU Compute
Department of Applied Mathematics and Computer Science

Reading Material

Reading material:
C17, C18
Feedback Groups of the day:

- Nathalie Jaure, Andine Havelange, Anne-Line Evenstad Dahlen
- Rebecca Christiansen, Line Karin Mortensen, Simon Eugen Matell
- Andreas Vedel Jantzen, Anne Sofie Talleruphuus
- Mathias Kirkeskov Madsen, Andreas Gramstrup Correia, Iulian Cozma
- Arnor Ingi Sigurdsson, Paul Jacques Connetable, Grigory Solomatov
- David Rosendahl Pedersen, Jacob Bundgaard Knudsen, Sune Mikkel Rasmussen
- Vilde Olerud, Rebecca Mølgaard Sommer,

Tue Herlau, Mikkel N. Schmidt and Morten Mørup

Introduction to Machine Learning and Data Mining

Course notes fall 2016, version 1

August 29, 2016

Technical University of Denmark

Lecture Schedule

1. Introduction
   30 August: C1

Data: Feature extraction, and visualization

2. Data and feature extraction
   6 September: C2, C3
3. Measures of similarity and summary statistics
   13 September: C4
4. Data Visualization and probability
   20 September: C5, C6

Supervised learning: Classification and regression

5. Decision trees and linear regression
   27 September: C7, C8 **(Project 1 due before 13:00)**
6. Overfitting and performance evaluation
   4 October: C9
7. Nearest Neighbor, Bayes and Naive Bayes
   11 October: C10, C11
8. Artificial Neural Networks and Bias/Variance
   25 October: C12, C13
9. AUC and ensemble methods
   1 November: C14, C15

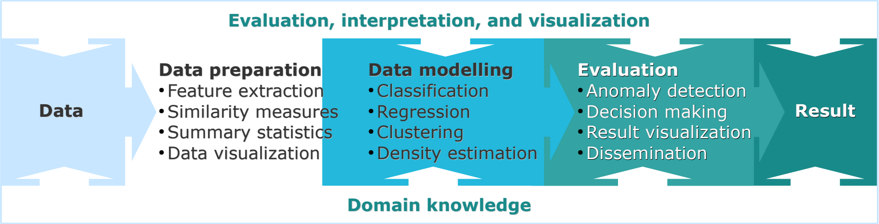
Unsupervised learning: Clustering and density estimation

10. K-means and hierarchical clustering
    8 November: C16 **(Project 2 due before 13:00)**
11. **Mixture models and density estimation**
    15 November: C17, C18
12. Association mining
    22 November: C19

Recap

13. Recap and discussion of the exam
    29 November: C1-C19 **(Project 3 due before 13:00)**

Data modeling framework



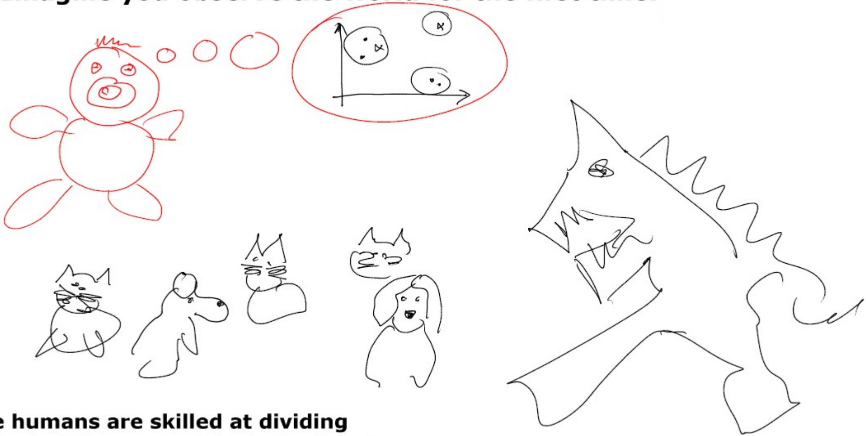
After today you should be able to:

Explain the role of the parameters in the Gaussian Mixture Model (GMM) and how the parameters are updated using the EM-algorithm

Explain why cross-validation can be used for GMM

Understand and apply kernel density, K-nearest neighbour density and average relative density estimation for outlier/anomaly detection.

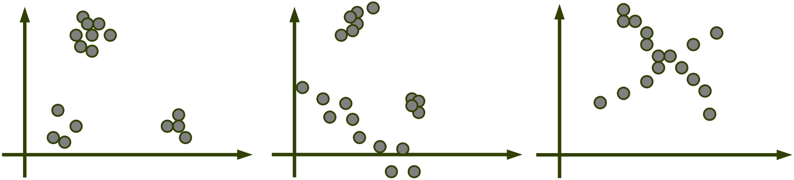
Imagine you observe the world for the first time!

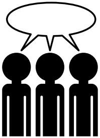


We humans are skilled at dividing objects into groups (clustering), but how do we make computers do the same?

Group exercise

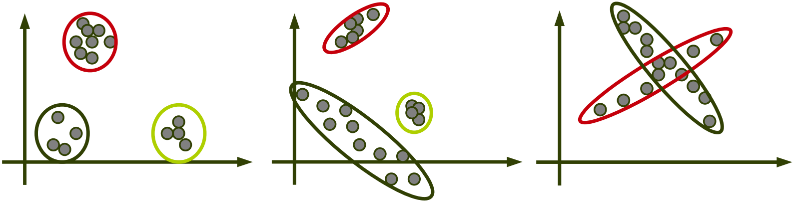
- What are the clusters below and what characterize each cluster?
- Is **k-means** well suited for modeling the clusters below?
  - Will it always find the optimum solution?
  - Can it model the sizes of the clusters?
  - Can it model the shape of the clusters?
  - How can we determine the number of clusters?





Group exercise

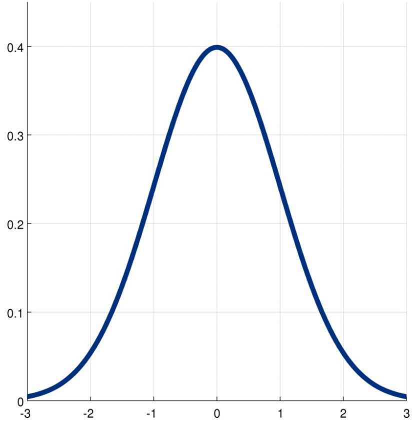
- What are the clusters below and what characterize each cluster?
- Is **k-means** well suited for modeling the clusters below?
  - Will it always find the optimum solution?
  - Can it model the sizes of the clusters?
  - Can it model the shape of the clusters?
  - How can we determine the number of clusters?



Normal distribution

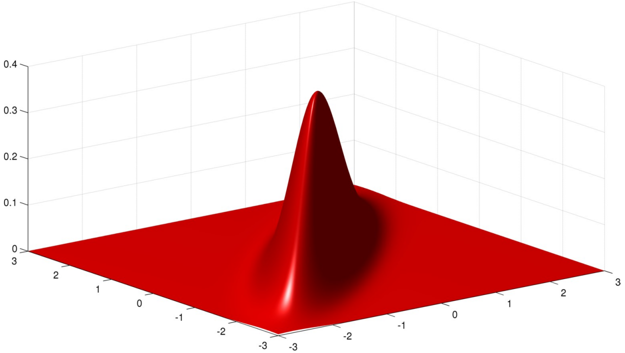
- Probability density function describes the relative chance of a given value to occur
- Normal distribution characterized by
  - Mean
  - Variance

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$



Multivariate Normal distribution

$$p(x) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \exp\left(-\tfrac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right)$$



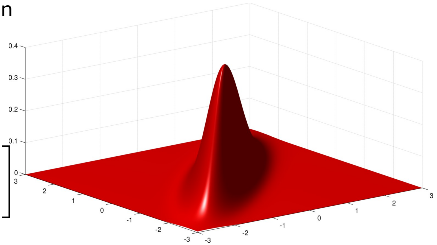
Multivariate Normal distribution

$$p(x) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right)$$

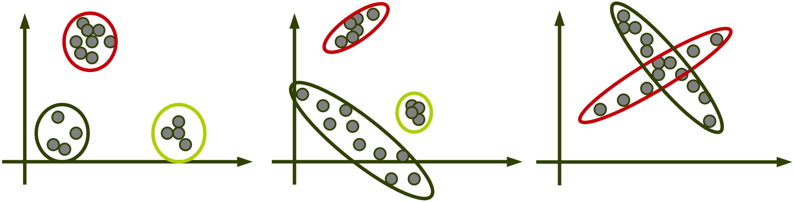
- Example: 2-dimensional Normal distribution

$$\mu = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \mathrm{var}(x_1) & \mathrm{cov}(x_1, x_2) \\ \mathrm{cov}(x_2, x_1) & \mathrm{var}(x_2) \end{bmatrix}$$



The Gaussian Mixture Model (GMM)



- Different locations $\boldsymbol{\mu}_{(k)}$
- Different shape $\boldsymbol{\Sigma}_{(k)}$
- Different sizes w_k

Data density

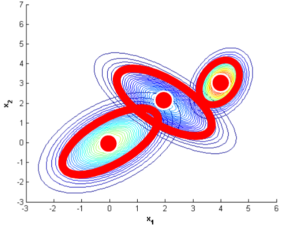
Sum of cluster specific densities assumed normal distributed

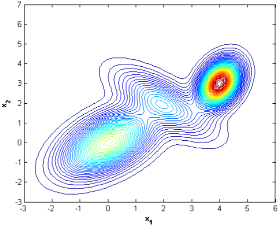
$$p(\mathbf{x}) = \sum_{k=1}^{K} w_k \mathcal{N}(\mathbf{x}, \mu_{(k)}, \boldsymbol{\Sigma}_{(k)})$$

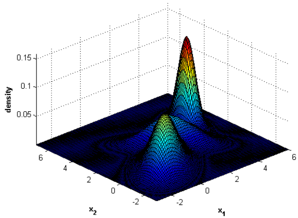
$$\text{s.t. } \sum_{k=1}^{K} w_k = 1,\ w_k \geq 0$$

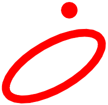
GMM example

$$p(\boldsymbol{x}) = 0.5\mathcal{N}\left(\boldsymbol{x}\middle|\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & 0.5\\0.5 & 1\end{bmatrix}\right) + 0.2\mathcal{N}\left(\boldsymbol{x}\middle|\begin{bmatrix}2\\2\end{bmatrix},\begin{bmatrix}1 & -0.7\\-0.7 & 1\end{bmatrix}\right) + 0.3\mathcal{N}\left(\boldsymbol{x}\middle|\begin{bmatrix}4\\3\end{bmatrix},\begin{bmatrix}0.2 & 0.1\\0.1 & 0.5\end{bmatrix}\right)$$









$\boldsymbol{\mu}_{(k)}$: Cluster center (prototypical example in cluster)

$\boldsymbol{\Sigma}_{(k)}$: Shape of the cluster

w_k : Relative size/density of the cluster

Group exercise

- Consider the Gaussian mixture model (GMM)

$$p(\mathbf{x}) = \sum_{k=1}^{K} w_k \mathcal{N}(\mathbf{x}|\,\mu_{(k)}, \mathbf{\Sigma}_{(k)}) \quad \text{s.t.} \ \sum_{k=1}^{K} w_k = 1,\ \ w_k \geq 0$$

- What is the value of the integral?

$$\int p(\mathbf{x}) \mathrm{d}\mathbf{x}$$

Gaussian mixture models, EM algorithm

Select an initial set of model parameters (mean and covariance for each cluster)

Repeat

- **Expectation**
  - For each object, calculate the probability of belonging to each distribution
- **Maximization**
  - For each probability distribution, estimate parameters by maximum likelihood

Until the parameters do not change

E-step

$$\ldots = k|\boldsymbol{x}_n) = \frac{w_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_{(k)}, \boldsymbol{\Sigma}_{(k)})}{\sum_{K=1}^{K} w_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_{(k)}, \boldsymbol{\Sigma}_{(k)})}$$

M-step

$$N_k = \sum_{n=1}^{N} p(\boldsymbol{z}_n = k|\boldsymbol{x}_n)$$

$$\boldsymbol{\mu}_{(k)} = \frac{1}{N_k} \sum_{n=1}^{N} \boldsymbol{x}_n p(\boldsymbol{z}_n = k|\boldsymbol{x}_n)$$

$$w_k = \frac{N_k}{N}$$

$$\boldsymbol{\Sigma}_{(k)} = \frac{1}{N_k} \sum_{n=1}^{N} (\boldsymbol{x}_n - \boldsymbol{\mu}_{(\boldsymbol{k})})(\boldsymbol{x}_n - \boldsymbol{\mu}_{(\boldsymbol{k})})^\top p(\boldsymbol{z}_n = k|\boldsymbol{x}_n)$$

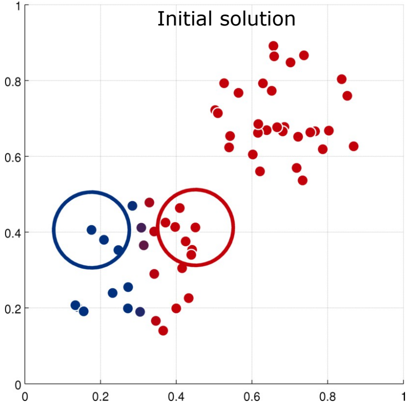
Gaussian mixture models, EM algorithm

Select an initial set of model parameters (mean and covariance for each cluster)

Repeat

- **Expectation**
  - For each object, calculate the probability of belonging to each distribution
- **Maximization**
  - For each probability distribution, estimate parameters by maximum likelihood

Until the parameters do not change



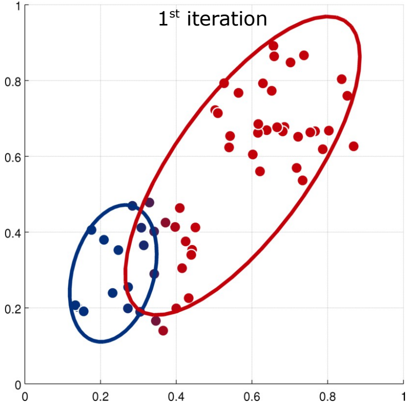
Gaussian mixture models, EM algorithm

Select an initial set of model parameters (mean and covariance for each cluster)

Repeat

- **Expectation**
  - For each object, calculate the probability of belonging to each distribution
- **Maximization**
  - For each probability distribution, estimate parameters by maximum likelihood

Until the parameters do not change



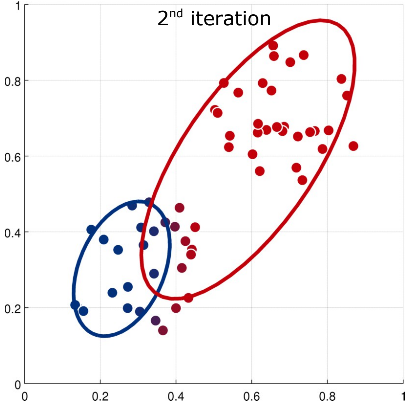
Gaussian mixture models, EM algorithm

Select an initial set of model parameters (mean and covariance for each cluster)

Repeat

- **Expectation**
  - For each object, calculate the probability of belonging to each distribution
- **Maximization**
  - For each probability distribution, estimate parameters by maximum likelihood

Until the parameters do not change



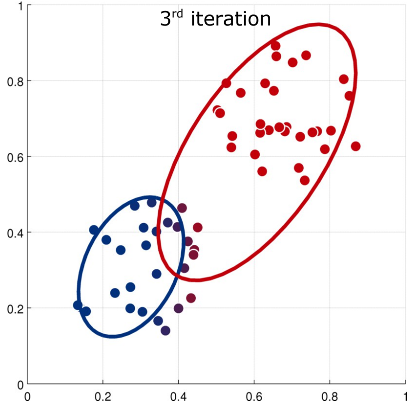
Gaussian mixture models, EM algorithm

Select an initial set of model parameters (mean and covariance for each cluster)

Repeat

- **Expectation**
  - For each object, calculate the probability of belonging to each distribution
- **Maximization**
  - For each probability distribution, estimate parameters by maximum likelihood

Until the parameters do not change



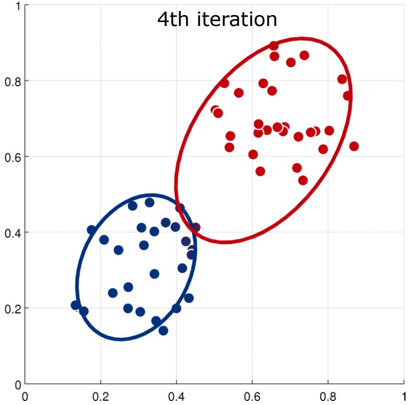
Gaussian mixture models, EM algorithm

Select an initial set of model parameters (mean and covariance for each cluster)

Repeat

- **Expectation**
  - For each object, calculate the probability of belonging to each distribution
- **Maximization**
  - For each probability distribution, estimate parameters by maximum likelihood

Until the parameters do not change



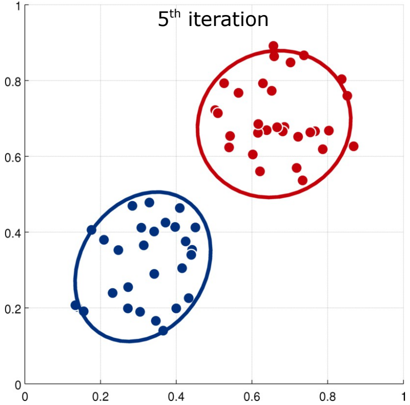
Gaussian mixture models, EM algorithm

Select an initial set of model parameters (mean and covariance for each cluster)

Repeat

- **Expectation**
  - For each object, calculate the probability of belonging to each distribution
- **Maximization**
  - For each probability distribution, estimate parameters by maximum likelihood

Until the parameters do not change



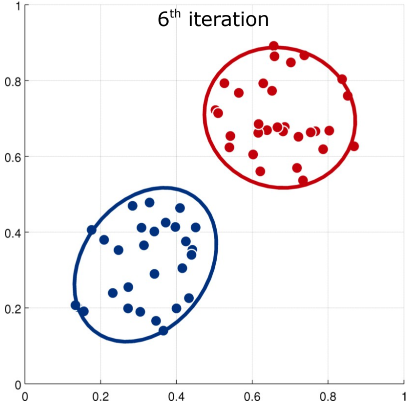
Gaussian mixture models, EM algorithm

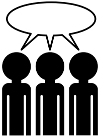
Select an initial set of model parameters (mean and covariance for each cluster)

Repeat

- **Expectation**
  - For each object, calculate the probability of belonging to each distribution
- **Maximization**
  - For each probability distribution, estimate parameters by maximum likelihood

Until the parameters do not change

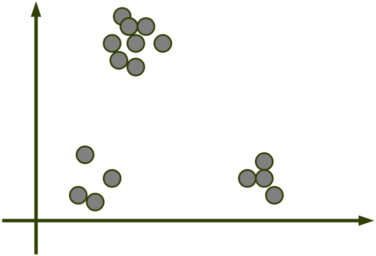


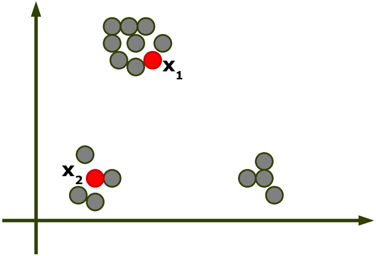


Group exercise

- Consider the data to the right with 16 observations.
  - What would ideally happen if we used a GMM with K=16 clusters to model the data?

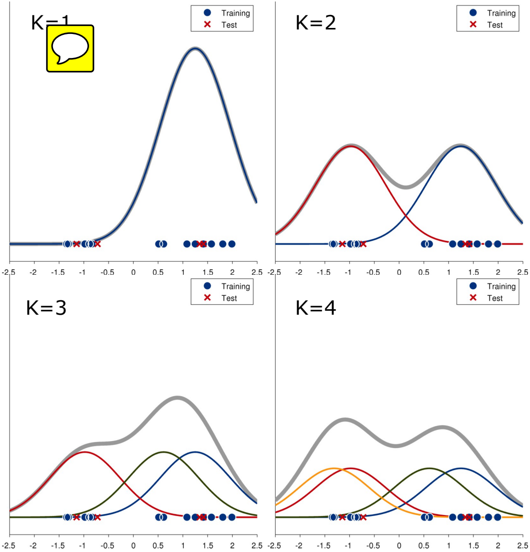
- Imagine we have two **test observations** denoted $\mathbf{x}_1$ and $\mathbf{x}_2$ (red points) that are not used for training.
  - What happens to p($\mathbf{x}_1$) and p($\mathbf{x}_2$) if we use K=3 and K=16 clusters





Mixture models

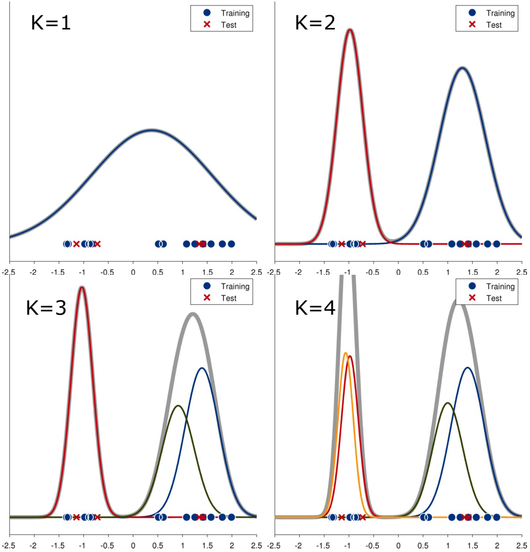
- Selecting complexity using crossvalidation



Mixture models

- Selecting complexity using crossvalidation

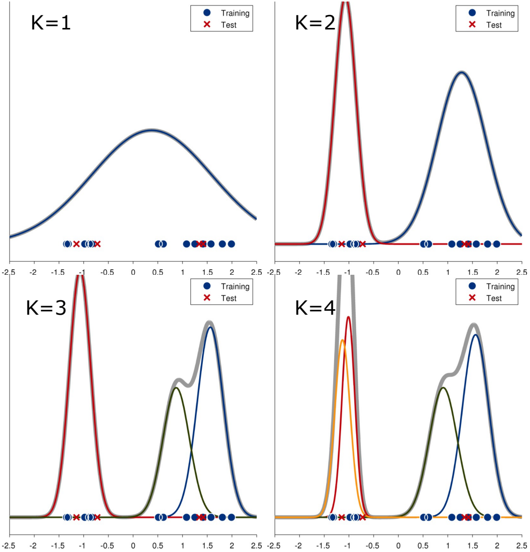
EM 1st iteration



Mixture models

- Selecting complexity using crossvalidation

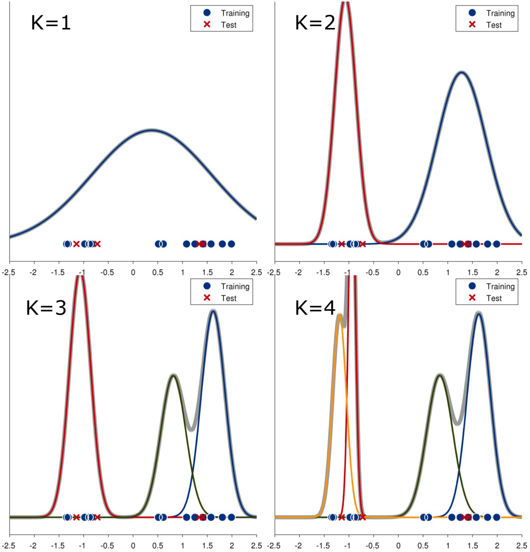
EM 2nd iteration



Mixture models

- Selecting complexity using crossvalidation

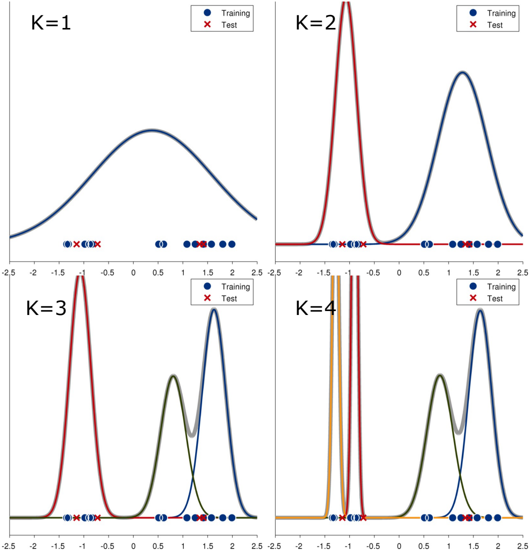
EM 3rd iteration



EM 4^{th} iteration

Mixture models

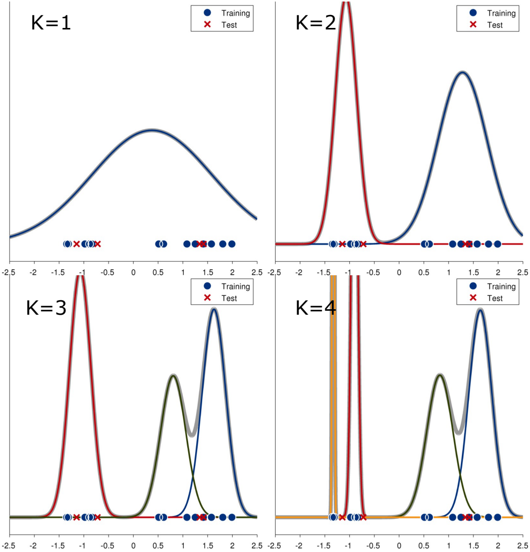
- Selecting complexity using crossvalidation



Mixture models

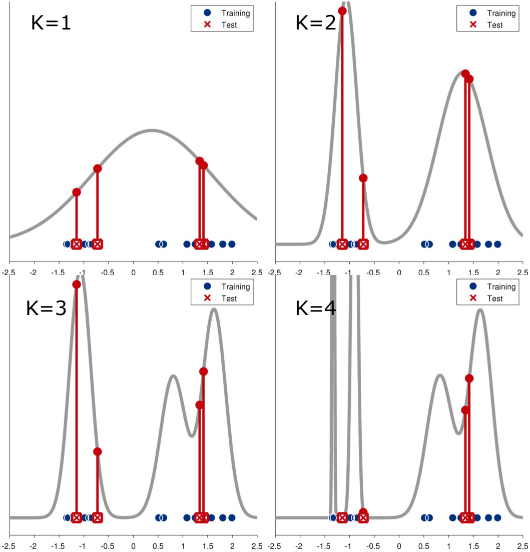
- Selecting complexity using crossvalidation

EM 5^{th} iteration



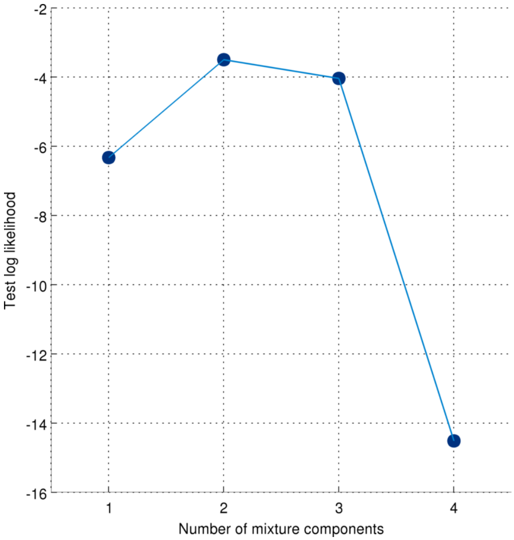
Mixture models

- Selecting complexity using crossvalidation



Mixture models

- Selecting complexity using crossvalidation



K-means versus GMM

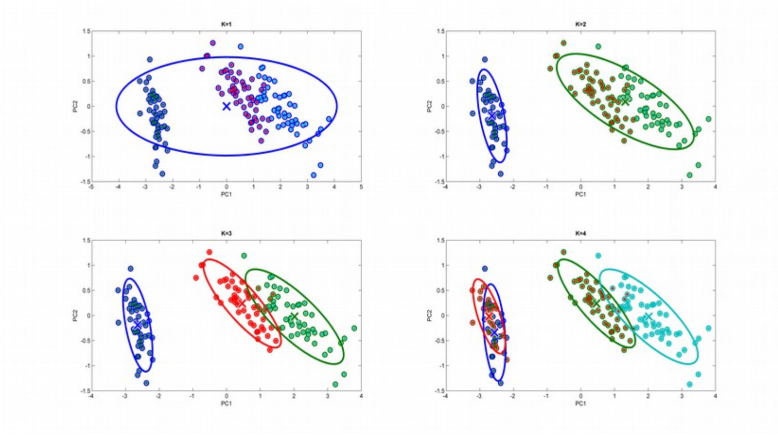
K-means

- No guarantee of optimal solution
- Does not model shape of clusters
- Does not model the size of clusters
- Difficult to assess the number of clusters to use particularly when there is no ground truth

Gaussian mixture model (GMM)

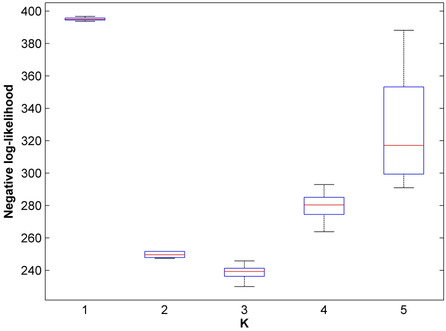
- No guarantee of optimal solution (even more local minima issues due to the additional model parameters)
- Models shape of cluster as ellipsoid
- Models the size of clusters
- Possible to estimate the number of components by cross-validation

GMM on Iris data using 1,2,3 and 4 components



Recap of GMM on Iris data

GMM 10 fold cross-validation on Iris data repeated five times where the five runs are plotted using box-plots.



Anomaly detection: Definition

- Given a collection of data objects
  - Each object has associated a number of features
- Detect which objects **deviate from normal** behaviour

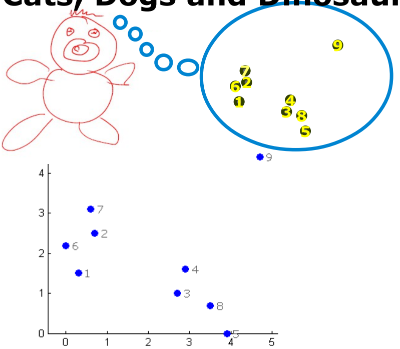
Anomaly detection: Example

- Credit card **fraud detection**
  - Recognize dubious credit card transactions based on the transaction history of the card holder
- Network **intrusion detection**
  - Detect hacker attacks, web crawlers etc.
- **Ecosystem disturbances**
  - Detect hurricanes, floods droughts, heat waves and fires
- **Health and medicine monitoring**
  - Detect abnormal behaviour in populations and patients
- **Fault detection in industry systems**
  - Detect when a wind turbine performs poorly due to ice coating on blades
- Detection of **outliers** in data measurements
  - Remove erroneous measurements due to misreading from an instrument

Group exercise

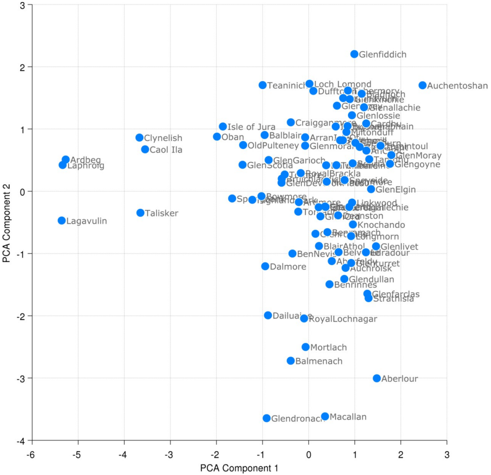
- Come up with **your own definition** of an outlier / anomaly
- How can we detect outliers using some of the methods you have already learned in the course?

Data example I: Cats, Dogs and Dinosaurs

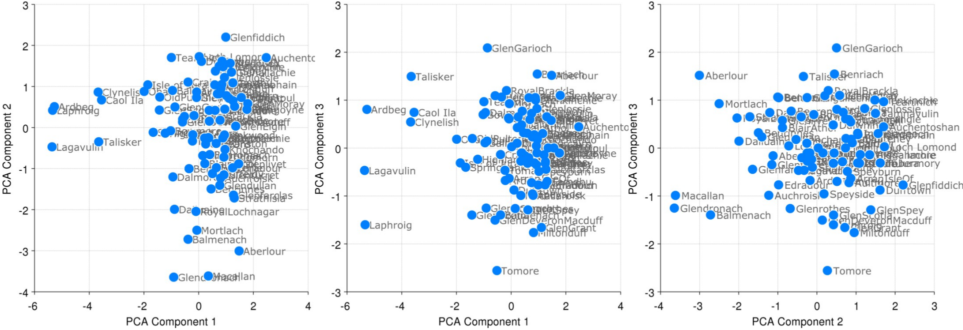


Data example II: Whisky

- 86 types of Scotch whisky
- Human ratings 1-5
- 12 taste categories
  - body, sweetness, smoky, medicinal, tobacco, honey, spicy, winey, nutty, malty, fruity, floral



PCA plot

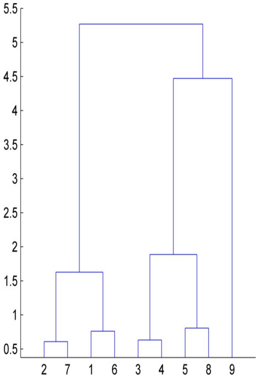


PCA Component 1
PCA Component 2
PCA Component 3
Glenfiddich
GlenGarioch
Talisker
Ardbeg
Laphroig
Lagavulin
Caol Ila
Clynelish
Mortlach
Balmenach
Aberlour
Macallan
Glendronach
Tomore
RoyalLochnagar
Auchroisk
Speyside
Glenrothes
GlenSpey
Miltonduff
Strathisla
Glendullan
Loch Lomond

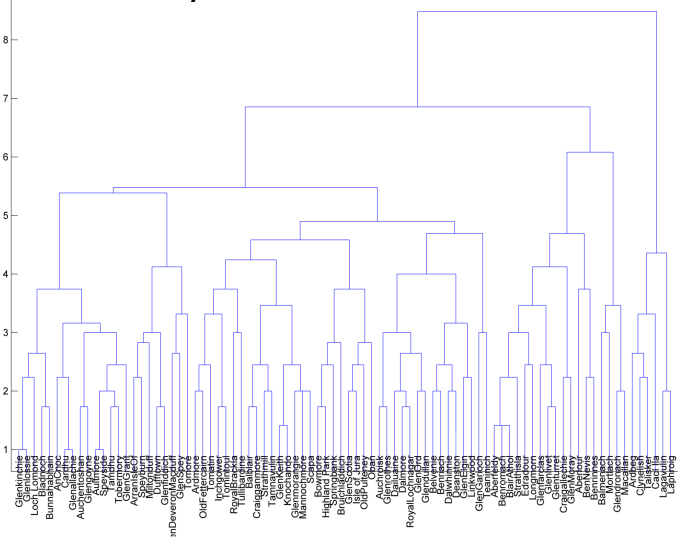
Dendrogram

- Dendrograms can be used to visualize relative distances between the observations

Data I: Cats, Dogs and Dinosaurs



Data II: Whisky



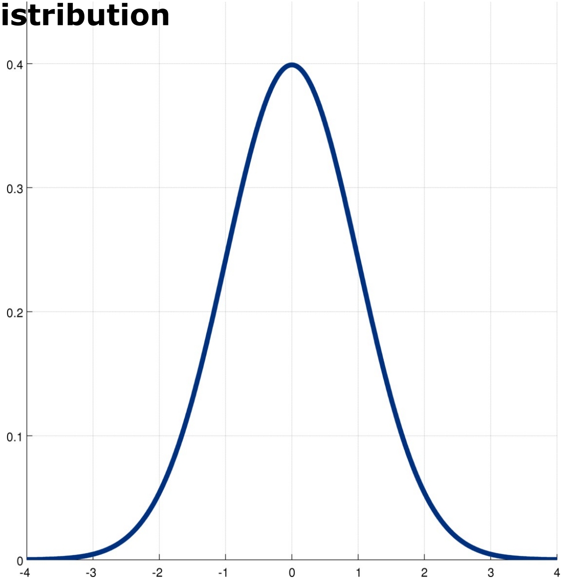
Density based techniques: Univariate normal distribution

- Map attribute to standard Normal variable

$$z = \frac{x - \mu}{\sigma}$$

- Choose a threshold

c	p(|z|>c)
1.0	0.3173
1.5	0.1336
2.0	0.0455
2.5	0.0124
3.0	0.0027
3.5	0.0005
4.0	0.0001



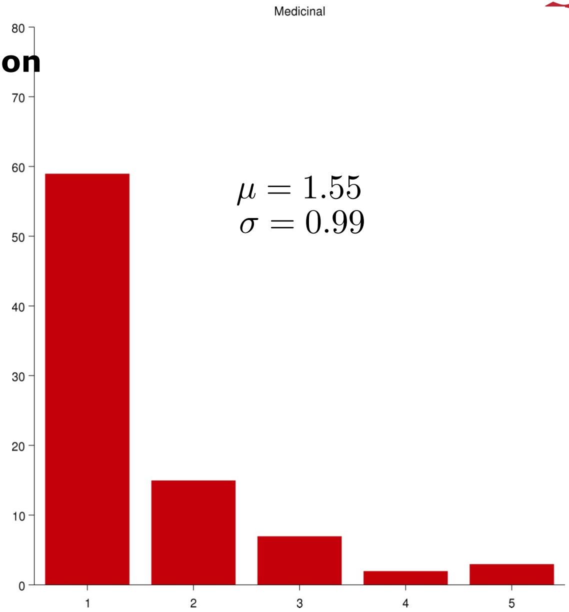
Normal distribution

- Map attribute to standard Normal variable

$$z = \frac{x-\mu}{\sigma}$$

- Choose a threshold

c	p(|z|>c)
1.0	0.3173
1.5	0.1336
2.0	0.0455
2.5	0.0124
3.0	0.0027
3.5	0.0005
4.0	0.0001



$$\mu = 1.55$$
$$\sigma = 0.99$$

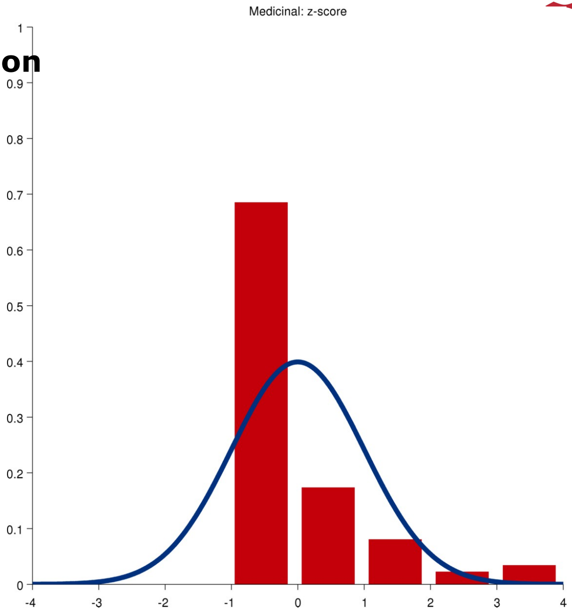
Normal distribution

- Map attribute to standard Normal variable

$$z = \frac{x - \mu}{\sigma}$$

- Choose a threshold

c	p(|z|>c)
1.0	0.3173
1.5	0.1336
2.0	0.0455
2.5	0.0124
3.0	0.0027
3.5	0.0005
4.0	0.0001



Normal distribution

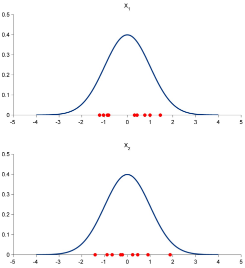
- Map attribute to standard Normal variable
- Choose a threshold

$$z = \frac{x-\mu}{\sigma}$$

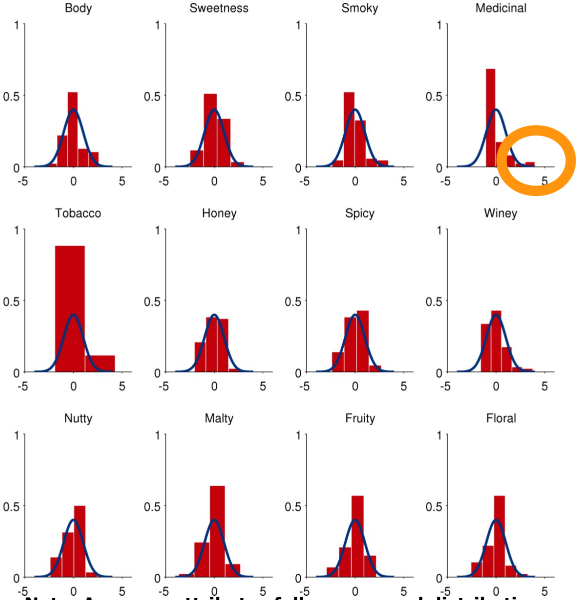
$$p(|z| > c) = 0.001$$

$$c = 3.2905$$

Data I: Cats, Dogs and Dinosaurs



Data II: Whisky



Note: Assumes attributes follow a normal distribution which may not be a valid assumption!

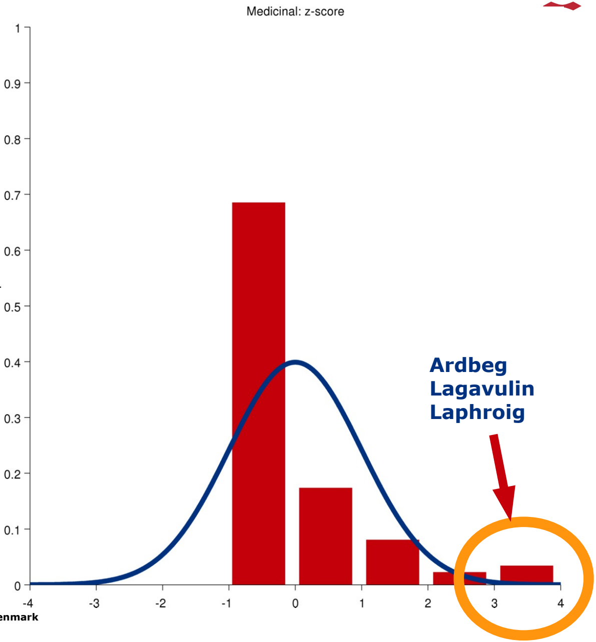
Normal distribution

- Map attribute to standard Normal variable

$$z = \frac{x - \mu}{\sigma}$$

- Choose a threshold

$$p(|z| > c) = 0.001$$
$$c = 3.2905$$



Approaches to anomaly detection

- **Density-based techniques**
  - Estimate the density of data objects
  - Outliers are:
    - Data objects in low density area

- **We can of course use the GMM to evaluate the density of test data.**
  - why not on the training data?

- **Approaches we will presently also consider**:
  - Kernel density estimation
  - Inverse average distance to K nearest neighbours (KNN density)
  - Average relative KNN density

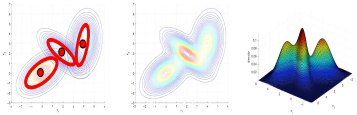
Density based techniques: Kernel Density Estimator

Rememer from Last week:
Gaussian Mixture Model (GMM)

Data density **Sum of cluster specific densities assumed normal distributed**

$$p(\boldsymbol{x}) = \sum_{K=1}^{K} w_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_{(k)}, \boldsymbol{\Sigma}_{(k)})$$

$$(s.t. \sum_{k=1}^{K} w_k = 1, \quad w_k \geq 0)$$

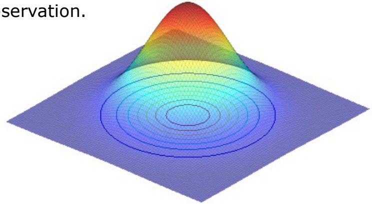


$\boldsymbol{\mu}_{(k)}$: Cluster center (prototypical example in cluster)

$\boldsymbol{\Sigma}_{(k)}$: Shape of the cluster

w_k : Relative density of the cluster

Kernel Density estimation based on Gaussian Kernel: Consider the GMM and define a Gaussian with mean $\boldsymbol{x}_n$ and co-variance $\sigma^2\boldsymbol{I}$ around each Observation.



Let all observation weight the same, i.e. $w_n=1/N$

$$p(\mathbf{x}) = \sum_{n=1}^{N} \frac{1}{N} \mathcal{N}(\mathbf{x}|\mathbf{x}_n, \sigma^2\mathbf{I})$$

Only free parameter σ^2!

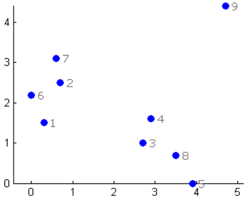
There is nothing special about the normal distribution. For a general mixture distribution p the general form of kernel density estimator is:

$$p(\mathbf{x}) = \sum_{n=1}^{N} \frac{1}{N} p(\mathbf{x}|\mathbf{x}_n, \theta)$$

This may be useful if x is discrete or non-negative.

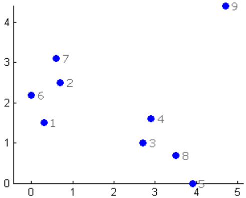
How do we determine σ^2?

Data I: Cats, Dogs and Dinosaurs

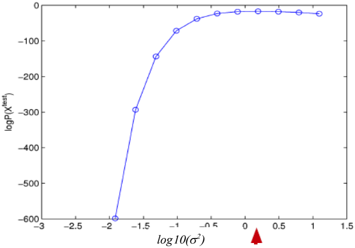


How do we determine σ^2

Data I: Cats, Dogs and Dinosaurs



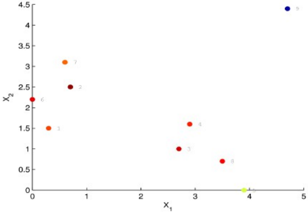
Density of test set based on leave-one-out cross validation

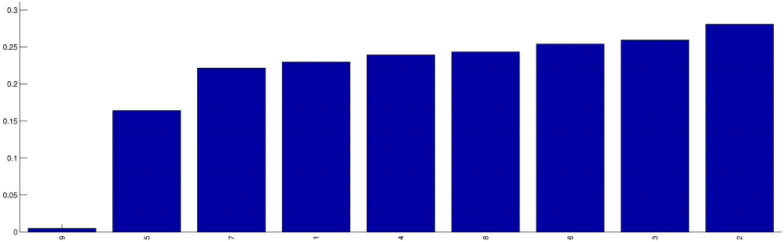


Optimal σ^2=1.55

σ^2=0.01 σ^2=0.1 σ^2=1 σ^2=5

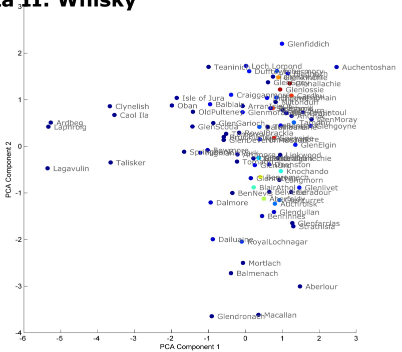
Etimated leave-one-out density evaluated at each observation



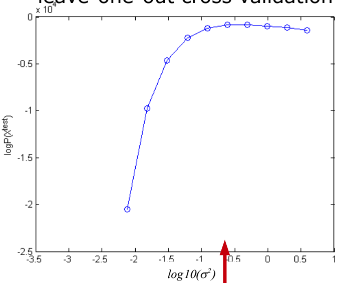


Etimated density evaluated at each observation

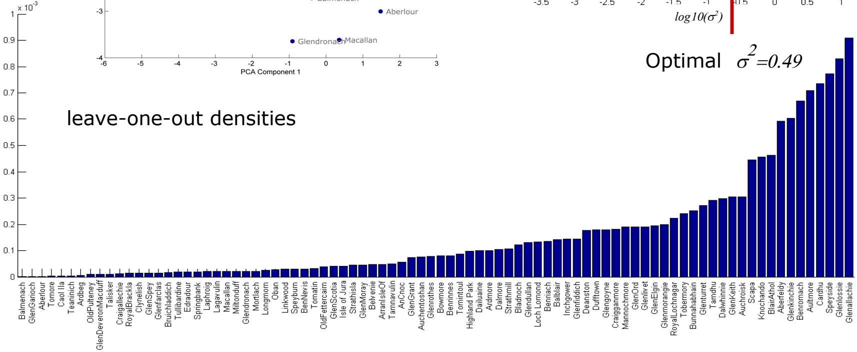
Data II: Whisky



Density of test set based on leave-one-out cross validation

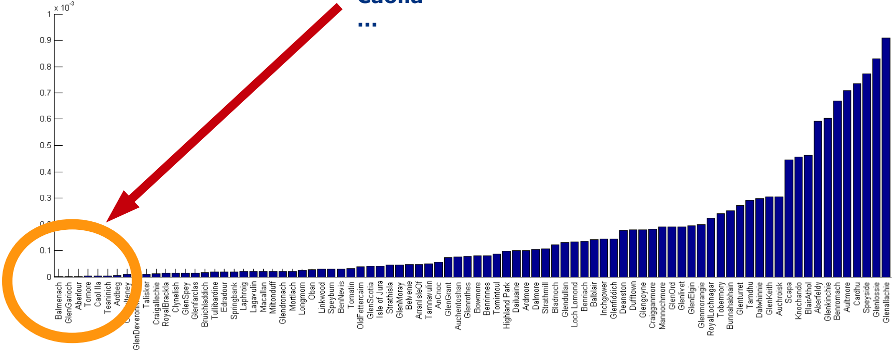


Optimal $\sigma^2=0.49$



Data II: Whisky

Balmenach
Glen Garioch
Aberlour
Tomore
Caolla
...



Inverse distance density estimation

- **Distance based measure of density**
  - Density is inverse proportional to avereage distance to k nearest neighbors
  - Density is low if nearest neighbors are far away

$$\text{density}_{\boldsymbol{X}_{\backslash i}}(\boldsymbol{x}_i, K) = \frac{1}{\frac{1}{K}\sum_{\boldsymbol{x}' \in N_{\boldsymbol{X}_{\backslash i}}(\boldsymbol{x}_i, K)} d(\boldsymbol{x}_i, \boldsymbol{x}')}$$

- **Relative density**
  - Density compared to density at nearest neighbors

$$\text{ard}_{\boldsymbol{X}}(\boldsymbol{x}_i, K) = \frac{\text{density}_{\boldsymbol{X}_{\backslash i}}(\boldsymbol{x}_i, K)}{\frac{1}{K}\sum_{\boldsymbol{x}_j \in N_{\boldsymbol{X}_{\backslash i}}(\boldsymbol{x}_i, K)} \text{density}_{\boldsymbol{X}_{\backslash j}}(\boldsymbol{x}_j, K)}$$

$$N_{\boldsymbol{X}}(\boldsymbol{x}, K) = \{\text{The } K \text{ observations in } \boldsymbol{X} \text{ which are nearest to } \boldsymbol{x}\}$$

$$\boldsymbol{X}_{\backslash i}^{T} = [\boldsymbol{x}_1 \ \boldsymbol{x}_2 \cdots \boldsymbol{x}_{i-2} \ \boldsymbol{x}_{i-1} \ \boldsymbol{x}_{i+1} \ \boldsymbol{x}_{i+2} \cdots \ \boldsymbol{x}_N]$$

These measures are taken from:
“Introduction to Data Mining” by Pang-Ning Tan, Michael Steinbach, Vipin Kumar

Consider the pairwise distance matrix given to the left. What is the density and average relative density of the first observation for k=2?

$d(\mathbf{x}_i,\mathbf{x}_j)$	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$	$\mathbf{x}_6$	$\mathbf{x}_7$
$\mathbf{x}_1$	0	2.5	2.4	4.0	0.8	0.6	3.3
$\mathbf{x}_2$	2.5	0	0.6	1.6	2.9	3.0	1.1
$\mathbf{x}_3$	2.4	0.6	0	1.9	3.0	2.7	1.0
$\mathbf{x}_4$	4.0	1.6	1.9	0	4.5	4.6	3.8
$\mathbf{x}_5$	0.8	2.9	3.0	4.5	0	1.1	3.9
$\mathbf{x}_6$	0.6	3.0	2.7	4.6	1.1	0	3.8
$\mathbf{x}_7$	3.3	1.1	1.0	3.8	3.9	3.8	0

$$\text{density}_{\boldsymbol{X}_{\setminus i}}(\boldsymbol{x}_i, K) = \frac{1}{\frac{1}{K}\sum_{\boldsymbol{x}' \in N_{\boldsymbol{X}_{\setminus i}}(\boldsymbol{x}_i, K)} d(\boldsymbol{x}_i, \boldsymbol{x}')}$$

$$\text{ard}_{\boldsymbol{X}}(\boldsymbol{x}_i, K) = \frac{\text{density}_{\boldsymbol{X}_{\setminus i}}(\boldsymbol{x}_i, K)}{\frac{1}{K}\sum_{\boldsymbol{x}_j \in N_{\boldsymbol{X}_{\setminus i}}(\boldsymbol{x}_i, K)} \text{density}_{\boldsymbol{X}_{\setminus j}}(\boldsymbol{x}_j, K)}$$

Consider the pairwise distance matrix given to the left. What is the density and average relative density of the first observation for k=2?

$d(\mathbf{x}_i,\mathbf{x}_j)$	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$	$\mathbf{x}_6$	$\mathbf{x}_7$
$\mathbf{x}_1$	0	2.5	2.4	4.0	0.8	0.6	3.3
$\mathbf{x}_2$	2.5	0	0.6	1.6	2.9	3.0	1.1
$\mathbf{x}_3$	2.4	0.6	0	1.9	3.0	2.7	1.0
$\mathbf{x}_4$	4.0	1.6	1.9	0	4.5	4.6	3.8
$\mathbf{x}_5$	0.8	2.9	3.0	4.5	0	1.1	3.9
$\mathbf{x}_6$	0.6	3.0	2.7	4.6	1.1	0	3.8
$\mathbf{x}_7$	3.3	1.1	1.0	3.8	3.9	3.8	0

$$\text{density}_{\boldsymbol{X}_{\setminus i}}(\boldsymbol{x}_i, K) = \frac{1}{\frac{1}{K}\sum_{\boldsymbol{x}' \in N_{\boldsymbol{X}_{\setminus i}}(\boldsymbol{x}_i, K)} d(\boldsymbol{x}_i, \boldsymbol{x}')}$$

$$\text{ard}_{\boldsymbol{X}}(\boldsymbol{x}_i, K) = \frac{\text{density}_{\boldsymbol{X}_{\setminus i}}(\boldsymbol{x}_i, K)}{\frac{1}{K}\sum_{\boldsymbol{x}_j \in N_{\boldsymbol{X}_{\setminus i}}(\boldsymbol{x}_i, K)} \text{density}_{\boldsymbol{X}_{\setminus j}}(\boldsymbol{x}_j, K)}$$

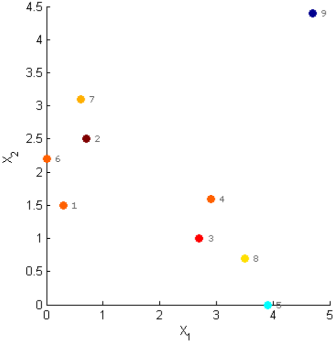
density($\mathbf{x}_1$,2)=$[1/2(0.6+0.8)]^{-1}$=2/1.4
density($\mathbf{x}_6$,2)=$[1/2(0.6+1.1)]^{-1}$=2/1.7
density($\mathbf{x}_5$,2)=$[1/2(0.8+1.1)]^{-1}$=2/1.9

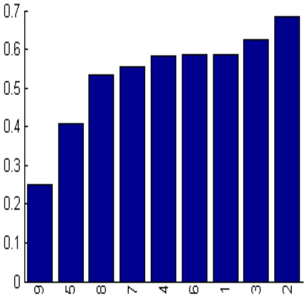
Av. Rel. Density($\mathbf{x}_1$,2)=[2/1.4]/[1/2 (2/1.7 + 2/1.9)]=1.2817

Inverse distance density estimation

- KNN density (5 nearest neighbors)

Data I: Cats , dogs and dinosaurs

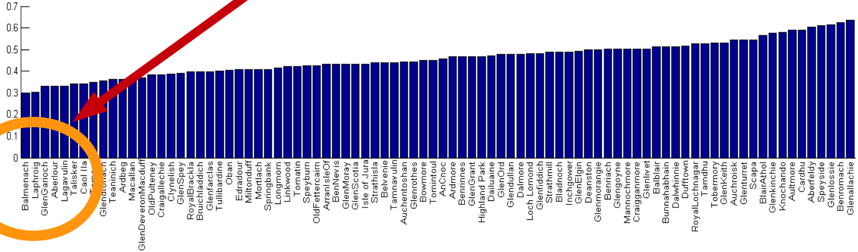




Inverse distance density estimation

- KNN density (5 nearest neighbors)

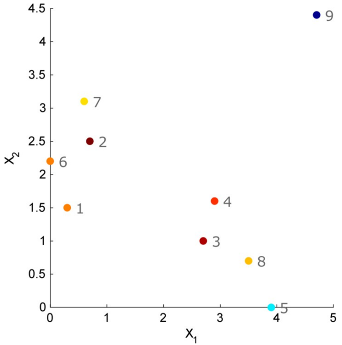
Balmenach
Laphroig
GlenGarioch
Aberlour
Lagevulin
...

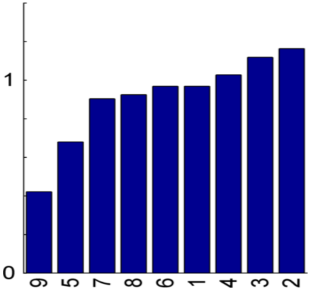


Average Relative density

- Average Relative KNN density (5 nearest neighbors)

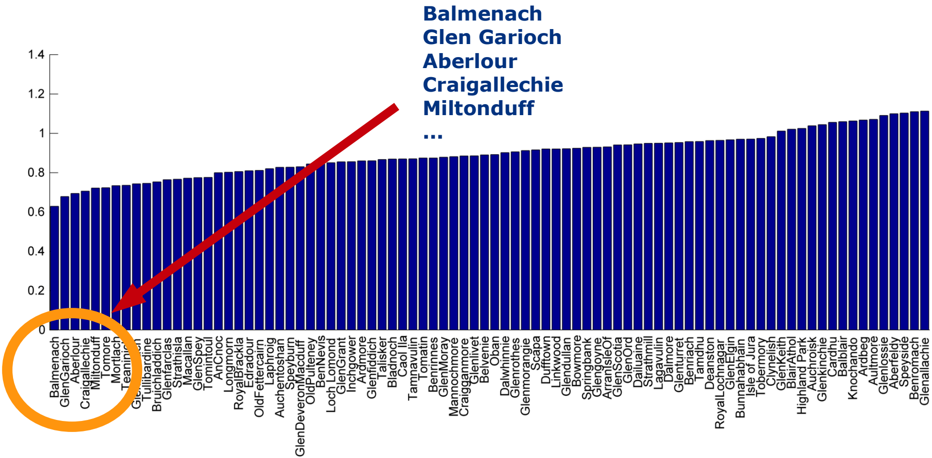
Data I: Cats , dogs and dinosaurs





Average relative density

- Average relative KNN density (5 nearest neighbors)



Results using different methods

- **Kernel Density Estimation**
  - Balmenach
  - Glen Garioch
  - Aberlour
  - Tomore
  - Caolla
- **KNN density**
  - Balmenach
  - Laphroig
  - Glen Garioch
  - Aberlour
  - Lagavulin
- **KNN average relative density**
  - Balmenach
  - Glen Garioch
  - Aberlour
  - Craigallechie
  - Miltonduff

  Common: Balmenach, Glen Garioch, Aberlour

Example of exam questions

QI: What is the average relative density for observation 2 (i.e. $\mathbf{x}_2$) for k=2 nearest neighbours?

A:1/5
B:3/10
C:7/10
D:1

$d(\mathbf{x}_i,\mathbf{x}_j)$	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$
$\mathbf{x}_1$	0	2.0	0.2	0.9	0.2
$\mathbf{x}_2$	2.0	0	1.5	0.5	2.0
$\mathbf{x}_3$	0.2	1.5	0	1.2	1.4
$\mathbf{x}_4$	0.9	0.5	1.2	0	1.0
$\mathbf{x}_5$	0.2	2.0	1.4	1.0	0

$$\text{density}_{\boldsymbol{X}_{\backslash i}}(\boldsymbol{x}_i, K) = \frac{1}{\frac{1}{K}\sum_{\boldsymbol{x}' \in N_{\boldsymbol{X}_{\backslash i}}(\boldsymbol{x}_i, K)} d(\boldsymbol{x}_i, \boldsymbol{x}')}$$

$$\text{ard}_{\boldsymbol{X}}(\boldsymbol{x}_i, K) = \frac{\text{density}_{\boldsymbol{X}_{\backslash i}}(\boldsymbol{x}_i, K)}{\frac{1}{K}\sum_{\boldsymbol{x}_j \in N_{\boldsymbol{X}_{\backslash i}}(\boldsymbol{x}_i, K)} \text{density}_{\boldsymbol{X}_{\backslash j}}(\boldsymbol{x}_j, K)}$$

Example of exam questions

QI: What is the average relative density for observation 2 (i.e. $\mathbf{x}_2$) for k=2 nearest neighbours?

A:1/5
B:3/10
C:7/10
D:1

$d(\mathbf{x}_i,\mathbf{x}_j)$	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$
$\mathbf{x}_1$	0	2.0	0.2	0.9	0.2
$\mathbf{x}_2$	2.0	0	1.5	0.5	2.0
$\mathbf{x}_3$	0.2	1.5	0	1.2	1.4
$\mathbf{x}_4$	0.9	0.5	1.2	0	1.0
$\mathbf{x}_5$	0.2	2.0	1.4	1.0	0

$$\text{density}_{\boldsymbol{X}_{\setminus i}}(\boldsymbol{x}_i, K) = \frac{1}{\frac{1}{K}\sum_{\boldsymbol{x}' \in N_{\boldsymbol{X}_{\setminus i}}(\boldsymbol{x}_i, K)} d(\boldsymbol{x}_i, \boldsymbol{x}')}$$

$$\text{ard}_{\boldsymbol{X}}(\boldsymbol{x}_i, K) = \frac{\text{density}_{\boldsymbol{X}_{\setminus i}}(\boldsymbol{x}_i, K)}{\frac{1}{K}\sum_{\boldsymbol{x}_j \in N_{\boldsymbol{X}_{\setminus i}}(\boldsymbol{x}_i, K)} \text{density}_{\boldsymbol{X}_{\setminus j}}(\boldsymbol{x}_j, K)}$$

QI: C