Dependence Modeling Using Copulas

Vaibhav Rajan
Research Scientist

Xerox Research Centre India

April 6, 2015

Measures of Dependence

- **Linear Correlation**

$$Corr = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)\, Var(X_2)}}$$

  - Zero correlation implies independence only for normal distribution
  - Invariant only to *linear* transformations,
    Example: log–normal RVs vs underlying normal RVs
  - Interval of attainable correlations shrink with some distributions: Small correlation does not imply small degree of dependence.

    Example: $X_1 \sim \mathcal{N}(0,1), X_2 = X_1^2$.
    $Cov(X_1, X_2) = E(X_1(X_1^2 - 1)) = E(X_1^3) - E(X_1) = 0$

Measures of Dependence

- **Rank Correlation**
  RVs X_1, X_2 with marginals F_1, F_2, Spearman's rho:

$$\rho_s = Corr(F_1(X_1), F_2(X_2))$$

  - CDF maps RV into a uniform variable
    Grade of X: $F_X(X) \sim \mathcal{U}_{[0,1]}$
  - Sample–based estimate: $Corr(R(X_1), R(X_2))$, R: Rank transformation
  - Measures monotonicity: ± 1 perfect monotonicity
- Other measures: Kendall's Tau, Mutual Information based measures...

Copulas

- Parametric models of dependence
- Can capture many different kinds of dependence, notably (a)symmetric tail dependence
- A p-dimensional copula is a multivariate distribution function (joint distribution of the grades)

$$C : [0,1]^p \mapsto [0,1]$$

$$C_U(u) = \mathcal{P}(F_{X_1}(X_1) \leq u_1, \ldots, F_{X_n}(X_n) \leq u_n)$$

Examples

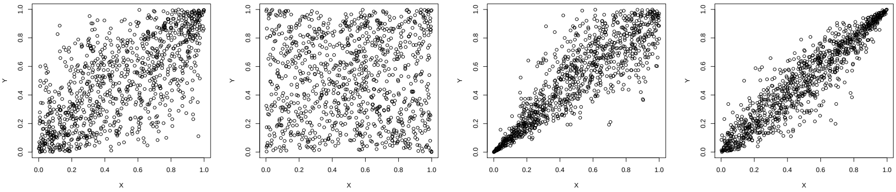


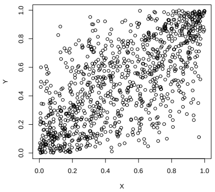
Figure: Gaussian, T, Clayton and Gumbel copulas (left to right).

Gaussian Copula

- For a given correlation matrix $R \in \mathbb{R}^{p \times p}$, the Gaussian copula is given by:

$$C_R^{\mathbf{N}}(u_1, \ldots, u_p) = \Phi_R\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_p)\right)$$

- Φ^{-1} is the inverse cumulative distribution function (CDF) of a standard Normal
- Φ_R is the joint cumulative distribution function of a multivariate Normal distribution with mean vector zero and covariance matrix equal to the correlation matrix R

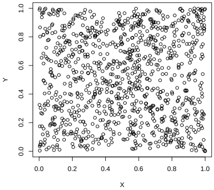


Student's t–copula

- For a correlation matrix Σ with ν degrees of freedom, the t–copula is defined as:

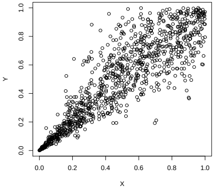
$$C_\theta(u_1, \ldots, u_p) = \mathbf{t}_{\nu,\Sigma}(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_p))$$

- $\mathbf{t}_{\nu,\Sigma}$: multivariate Student's t distribution
  t_ν: univariate t distribution
- Allows for joint fat tails and increased probability of joint extreme events (compared to Gaussian copula)
- Additional parameter ν: increasing ν decreases the tendency to exhibit extreme co–movements.



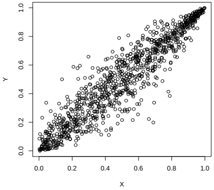
Clayton Copula

- $C_\delta(u,v) = (u^{-\delta} + v^{-\delta} - 1)^{-1/\delta}$
- δ controls the dependence.
  perfect dependence: $\delta \longrightarrow \infty$, independence: $\delta \longrightarrow 0$
- Asymmetric: greater dependence in negative tail



Gumbel Copula

- $C_\delta(u,v) = exp(-[(-\log u)^\delta + (-\log v)^\delta]^{1/\delta})$
- δ controls the dependence.
  perfect dependence: $\delta \longrightarrow \infty$, independence: $\delta = 1$
- Assymetric: greater dependence in the positive tail



Sklar's Theorem

Theorem

Let F be a joint distribution function with marginals $F_1, \ldots, F_p$. Then there exists a copula $C : [0, 1]^p \mapsto [0, 1]$ such that

$$F(x_1, \ldots, x_p) = C(F_1(x_1), \ldots, F_p(x_p))$$

If the marginal distributions are continuous, then this copula is unique. Conversely, if C is a copula and $F_1, \ldots, F_p$ are univariate distribution functions, then F as defined above is a multivariate distribution function with marginals $F_1, \ldots, F_p$.

- Model marginals and copula independently
- No constraints on marginals

Multivariate PDF

- Joint density: $f(x_1,\ldots,x_p) = C(F_1(x_1),\ldots,F_p(x_p))\ \Pi_{k=1}^{p} f_k(x_k)$
  - Absolutely continuous joint distribution F
  - Strictly increasing and continuous marginal distributions $F_1,\ldots,F_p$ and densities $f_1,\ldots,f_p$
- The copula density can be expressed in terms of the joint density and the marginals:

$$C(F_1(x_1),\ldots,F_p(x_p)) = \frac{f(x_1,\ldots,x_p)}{\Pi_{k=1}^{p} f_k(x_k)}.$$

Parameter Estimation

Many different methods exist
Inference from Margins:

- Estimate marginals (e.g. maximum likelihood fit)
- Assume copula family, c
- Maximum likelihood estimate of copula parameter θ

$$l(\theta) = \Sigma_1^n \log \left[c_\theta(\hat{F}_1(X_1), \ldots, \hat{F}_n(X_n))\right]$$

- Since copulas are invariant to monotonic transformations, a semi–parametric approach:

$$l(\theta) = \Sigma_1^n \log \left[c_\theta(R(X_1), \ldots, R(X_n))\right]$$

Parameter estimation algorithms can be improved

Gaussian Mixture Copula Model (GMCM)

- Tewari et al (2011) propose a Gaussian Mixture Copula Model:

$$\mathcal{C}(u_{i1}, u_{i2}, \dots, u_{ip} \mid \boldsymbol{\vartheta}) = \frac{\sum_{g=1}^{G} \pi_g \phi(\mathbf{y}_i \mid \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)}{\prod_{j=1}^{p} \psi_j(y_{ij})}$$

  $y_{ij} = \Psi_j^{-1}(u_{ij})$: inverse distribution values
  ψ_j: marginal density of the GMM along the j-th dimension
  Ψ_j^{-1}: inverse distribution function of the GMM along the j-th dimension
- Dependence structure from GMM
- No distributional assumption on the marginal is required
- Gradient–descent based heuristic: maximum–likelihood estimate of $\boldsymbol{\vartheta}$
- Application: clustering

Gaussian Mixture Copula Model (GMCM)

- For a heterogenous dataset $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)$,

$$\log \mathcal{L}\left(\boldsymbol{\vartheta} \mid \mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\right) = \sum_{i=1}^{n} \log \frac{\sum_{g=1}^{G} \pi_g \phi(\mathbf{y}_i \mid \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)}{\prod_{j=1}^{p} \psi_j(y_{ij})}$$

  - $\pi_g > 0$, with $\sum_{g=1}^{G} \pi_g = 1$, are mixing proportions,
  - $\phi(\mathbf{x} \mid \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ is multivariate Gaussian density with mean $\boldsymbol{\mu}_g$ and covariance matrix $\boldsymbol{\Sigma}_g$,
  - $\boldsymbol{\vartheta} = (\pi_1, \ldots, \pi_G, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_G, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_G)$,
  - $u_{ij} = F_j(x_{ij})$ and F_j is the unknown marginal distribution for the j-th dimension,
  - $y_{ij} = \Psi_j^{-1}(u_{ij})$ where Ψ_j and ψ_j are the marginal CDF and marginal pdf of the GMM along the j-th dimension.

EM for GMM

Initialize $\mu_g, \Sigma_g, \pi_g > 0, \qquad g = 1, \ldots, G.$
REPEAT

- **[E step]**

$$z_{ig}^{(t)} = \frac{\pi_g^{(t)} \phi(\mathbf{x}_i \mid \boldsymbol{\mu}_g^{(t)}, \boldsymbol{\Sigma}_g^{(t)})}{\sum_{g=1}^{G} \pi_g^{(t)} \phi(\mathbf{x}_i \mid \boldsymbol{\mu}_g^{(t)}, \boldsymbol{\Sigma}_g^{(t)}}$$

- **[M step]** For each g, update $\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g, \pi_g$ that maximizes likelihood:
  - $\pi_g^{(t+1)} = \frac{\sum_{i=1}^{n} z_{ig}^{(t)}}{n}$
  - $\boldsymbol{\mu}_g^{(t+1)} = \frac{\sum_{i=1}^{n} z_{ig}^{(t)} \mathbf{x}_i}{\sum_{i=1}^{n} z_{ig}^{(t)}}$
  - $\boldsymbol{\Sigma}_g^{(t+1)} = \frac{\sum_{i=1}^{n} z_{ig}^{(t+1)} \left(\mathbf{x}_i - \boldsymbol{\mu}_g^{(t+1)}\right)^T \left(\mathbf{x}_i - \boldsymbol{\mu}_g^{(t+1)}\right)}{\sum_{i=1}^{n} z_{ig}^{(t)}}$
  - Likelihood,
    $L^{(t+1)} = \prod_{i=1}^{n} \sum_{g=1}^{G} \pi_g^{(t+1)} \frac{1}{\sqrt{det(2\pi \boldsymbol{\Sigma}_g^{(t+1)})}} \times$
    $\exp\{-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu}_g^{(t+1)})^T \boldsymbol{\Sigma}_g^{(t+1)^{-1}} (\mathbf{x}_i - \boldsymbol{\mu}_g^{(t+1)})\}$

UNTIL termination criterion reached

A modified EM algorithm

- The y_{ij} values change at every iteration
- The conventional EM algorithm of a GMM, where the inputs remain fixed in each iteration, cannot be directly used here
- Our algorithm runs iteratively with the standard Expectation (E) and Maximization (M) steps and an additional step in each iteration
- The (E) and (M) steps update the parameters based on y_{ij}
- The additional step updates y_{ij} based on the parameters
- y_{ij} is truncated above to ensure convergence

A modified EM algorithm

- **Initialize:** Standardize the matrix $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p)$.
  Set $\boldsymbol{\vartheta}^{(0)}$ randomly or by K-means clustering under the constraints that $\pi_g^{(0)} > 0$, $\sum_{g=1}^{G} \pi_g^{(0)} = 1$ and $\boldsymbol{\Sigma}_g^{(0)}$ is positive definite, and set $\delta_i = \underset{g,j}{\text{Min}} \mid y_{ij}^{(0)} - 2\kappa^{(0)} \left(\left[\boldsymbol{\Sigma}_g^{(0)} + I \right]^{-1} \boldsymbol{\Sigma}_g^{(0)} \mathbf{1} \right)_j \mid$. Set $u_{ij} = F_j(x_{ij})$.
- **Repeat** the following steps **until** $|| L^{(t+1)} - L^{(t)} ||< \gamma$.
  - Set $y_{ij}^{(t)} = \min\left(\Lambda_{ij}^{(t)}, \Gamma_{ij}^{(t)}\right)$ where $\Lambda_{ij}^{(t)} = \left(\sum_{g=1}^{G} \frac{\pi_g^{(t)}}{\sigma_{g,jj}^{(t)}} \right)^{-1} \left[u_{ij} + \frac{1}{\sqrt{2\pi}} \sum_{g=1}^{G} \frac{\pi_g^{(t)} \mu_{gj}^{(t)}}{\sigma_{g,jj}^{(t)}} - \frac{1}{2} \right]$,
    $\Gamma_{ij}^{(t)} = \kappa^{(t)} \left(\left[S_i^{(t)} + I \right]^{-1} S_i^{(t)} \mathbf{1} \right)_j - \frac{\delta_i}{2} (3 - \frac{p}{m_i^{(t)} + p})$, $\kappa^{(t)} = \max_{g,j}(\mu_{gj}^{(t)})$, $S_i^{(t)} = \sum_{g=1}^{G} z_{ig}^{(t-1)} \boldsymbol{\Sigma}_g^{(t)}$, and $m_i^{(t)}$ is the sum of all elements of $S_i^{(t)}$.
  - **E-Step:** $z_{ig}^{(t)} = \frac{\pi_g^{(t)} \phi(\mathbf{y}_i^{(t)} | \boldsymbol{\mu}_g^{(t)}, \boldsymbol{\Sigma}_g^{(t)})}{\sum_{g=1}^{G} \pi_g^{(t)} \phi(\mathbf{y}_i^{(t)} | \boldsymbol{\mu}_g^{(t)}, \boldsymbol{\Sigma}_g^{(t)})}$.
  - **M-Step:**
    $$\pi_g^{(t+1)} = \frac{\sum_{i=1}^{n} z_{ig}^{(t)}}{n}, \boldsymbol{\mu}_g^{(t+1)} = \frac{\sum_{i=1}^{n} z_{ig}^{(t)} \mathbf{y}_i^{(t)}}{\sum_{i=1}^{n} z_{ig}^{(t)}}, \boldsymbol{\Sigma}_g^{(t+1)} = \frac{\sum_{i=1}^{n} z_{ig}^{(t+1)} \left(\mathbf{y}_i^{(t)} - \boldsymbol{\mu}_g^{(t+1)}\right)^T \left(\mathbf{y}_i^{(t)} - \boldsymbol{\mu}_g^{(t+1)}\right)}{\sum_{i=1}^{n} z_{ig}^{(t)}}$$
  - **Likelihood:**
    $$\mathcal{L}^{(t+1)} = \prod_{i=1}^{n} \sum_{g=1}^{G} \pi_g^{(t+1)} \frac{1}{\sqrt{det(2\pi \boldsymbol{\Sigma}_g^{(t+1)})}} \times \exp -\frac{1}{2} (\mathbf{y}_i^{(t)} - \boldsymbol{\mu}_g^{(t+1)})^T \boldsymbol{\Sigma}_g^{(t+1)\,-1} (\mathbf{y}_i^{(t)} - \boldsymbol{\mu}_g^{(t+1)})$$

Theoretical results

Lemma

For each i and j, with probability more than 0.9975, y_{ij} can be approximated as

$$y_{ij} \approx \left(\sum_{g=1}^{G} \frac{\pi_g}{\sqrt{\sigma_{g,jj}}} \right)^{-1} \left[u_{ij} + \frac{1}{\sqrt{2\pi}} \sum_{g=1}^{G} \frac{\pi_g \mu_{gj}}{\sqrt{\sigma_{g,jj}}} - \frac{1}{2} \right].$$

Lemma

If $\underset{j}{Max} \mid y_{ij}^{(t+1)} - y_{ij}^{(t)} \mid \leq \delta_i$, *then*

$$\log \mathcal{C}(\boldsymbol{\vartheta}^{(t+1)}, \mathbf{y}_i^{(t+1)} \mid \mathbf{z}_i^{(t)}) \geq \log \mathcal{C}(\boldsymbol{\vartheta}^{(t+1)}, \mathbf{y}_i^{(t)} \mid \mathbf{z}_i^{(t)}).$$

Theorem

There exists t_0 *such that* $\mathcal{L}(\boldsymbol{\vartheta}^{(t+1)} \mid \mathbf{u}_1, \ldots, \mathbf{u}_n) \geq \mathcal{L}(\boldsymbol{\vartheta}^{(t)} \mid \mathbf{u}_1, \ldots, \mathbf{u}_n)$ *for* $t \geq t_0$.

PGMM

- Large number of parameters in the covariance matrices
- Constraints are imposed as $\mathbf{\Sigma}_g = \mathbf{\Lambda}_g\mathbf{\Lambda}_g' + \mathbf{\Psi}$ where $\mathbf{\Psi}$ is the matrix of white noise, $\mathbf{\Lambda}$ is a p x q matrix of factor loadings and q is the number of latent factors
- Mcnicholas and Murphy (2008) put constraints on the loading and noise structures to derive eight parsimonious covariance structures

Model ID	$\mathbf{\Lambda}_g$	$\mathbf{\Psi}_g$	$\mathbf{\Psi}_g = \psi_g\mathbf{I}$	Cov. Parameters
CCC	Constrained	Constrained	Constrained	$[pq - q(q-1)/2] + 1$
CCU	Constrained	Constrained	Unconstrained	$[pq - q(q-1)/2] + p$
CUC	Constrained	Unconstrained	Constrained	$[pq - q(q-1)/2] + G$
CUU	Constrained	Unconstrained	Unconstrained	$[pq - q(q-1)/2] + Gp$
UCC	Unconstrained	Constrained	Constrained	$G[pq - q(q-1)/2] + 1$
UCU	Unconstrained	Constrained	Unconstrained	$G[pq - q(q-1)/2] + p$
UUC	Unconstrained	Unconstrained	Constrained	$G[pq - q(q-1)/2] + G$
UUU	Unconstrained	Unconstrained	Unconstrained	$G\{pq - q(q-1)/2\} + Gp$

The BIC

- The BIC (Schwartz, 1978) is used to select the best member of the PGMM family
- The BIC can be written

$$\text{BIC} = 2l(\mathbf{x}, \hat{\boldsymbol{\vartheta}}) - \rho \log n,$$

where $\hat{\boldsymbol{\vartheta}}$ is the MLE of $\boldsymbol{\vartheta}$, ρ is the number of free parameters and n is the number of observations
- Keribin (1998, 2000) shows that the BIC gives consistent estimates of the number of mixture components under certain regularity conditions

Simulated data

- We generate 25 simulations each for eight different simulation settings
- Each simulation setting has 2-dimensional data points in four clusters
- Each data point is a product of a sample from a Multivariate Normal (MVN) distribution and another distribution as outlined in the table below
- In each simulation we generate the clusters by sampling from the distributions, f_i
- The tables in the next slide show the cluster distributions and sizes for each simulation setting

Simulation setup

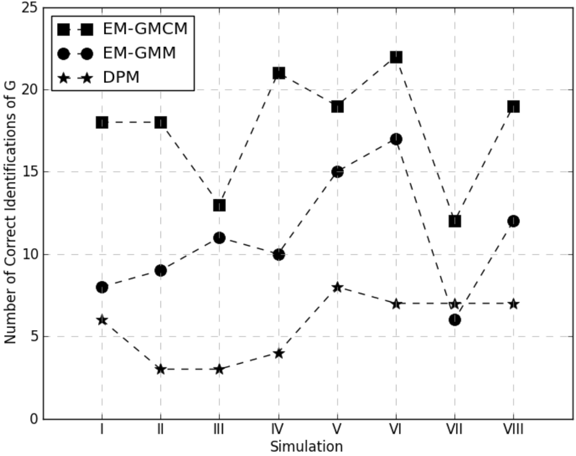
$f_1 = \mathrm{MVN}(-5.5, I_2/2) \times \mathrm{Unif}(0, 1)$
$f_2 = \mathrm{MVN}(2, D_2) \times \mathrm{t}(\mathrm{df} = 9)$
$f_3 = \mathrm{MVN}(3, \Delta) \times \mathrm{C}(\mathrm{loc} = 0, \mathrm{sc} = 1)$
$f_4 = \mathrm{MVN}(-5.5, I_2/2) \times \Gamma(\mathrm{sh} = 0.5, \mathrm{rt} = 1)$

Table: Parameter settings in each cluster; f_i: distribution in cluster $i = 1, 2, 3, 4$, I_2: 2×2 Identity matrix with, D_2: 2×2 diagonal matrix with unequal diagonal elements and Δ: matrix with (i, j)-th element $0.9^{|i-j|}$. Abbreviations – C: Cauchy, loc: location, sc: scale, sh: shape, rt: rate.

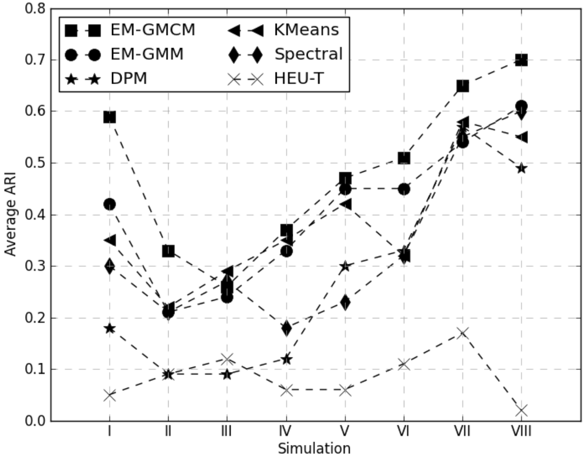
Sim	$\lvert c_1 \rvert$	$\lvert c_2 \rvert$	$\lvert c_3 \rvert$	$\lvert c_4 \rvert$	Data size
I	400	300	300	500	1500
II	300	500	300	400	1500
III	300	300	500	400	1500
IV	500	400	300	300	1500
V	1200	900	900	1500	4500
VI	900	1500	900	1200	4500
VII	900	900	1500	1200	4500
VIII	1500	1200	900	900	4500

Table: Parameter settings for each simulation set: $|c_i|$ denotes size of cluster c_i for $i = 1, 2, 3, 4$.

Identification of right number of clusters



Classification Accuracy



Real data experiment I

- Cleveland Heart data set, obtained from UCI Repository.
- 297 patients.
- We extract five numerical features–age, resting blood pressure, serum cholesterol, maximum heart rate, ST depression induced by exercise.
- The task is to classify the individuals into two groups: those with and those without heart disease.
- we set the range of G (the number of clusters) and q (number of latent factors) in our algorithms to $(1, 2, 3)$ and $(1, 2)$ respectively.
- For algorithms which do not estimate the number of clusters, such as K-means, Spectral clustering and HEU-T, we provide the correct number of clusters (2) as input.

Real data experiment I

- EM-GMCM detects the correct number of clusters, $G = 2$, with the CCU dependency structure and $q = 1$.
- EM-GMM incorrectly predicts 3 clusters with UCU dependence structure and $q = 2$.

Algorithm	EM-GMCM	EM-GMM	K-Means	Spectral	HEU-T
Accuracy	**70%**	15%	57%	35%	50%

EM-GMCM

	A	*B*
A	**137**	23
B	67	**70**

EM-GMM

	A	*B*	*C*
A	19	69	72
B	12	26	99

K-Means

	A	*B*
A	110	50
B	76	61

Spectral

	A	*B*
A	64	96
B	98	39

HEU-T

	A	*B*
A	114	46
B	103	34

Table: Above: Classification accuracy of the algorithms tested on the Cleveland Heart Disease Data. Below: Classification tables of the best models chosen by EM-GMCM and EM-GMM and of K-means, Spectral clustering and HEU-T for the Cleveland Heart data. *A*: group without heart disease, *B*: group with heart disease. EM-GMM erroneously estimates a third group *C*. True labels: horizontal, Estimated labels: vertical.

Real data experiment II

- Gamma Telescope data, obtained from UCI Repository.
- consists of features extracted from a preprocessed image of reconstructed radiation showers.
- 2000 observations with 10 features.
- The task is to statistically discriminate between two signals within the image: the primary gamma signal in Cherenkov radiation and background hadronic shower signal from cosmic rays in the upper atmosphere.
- We set the range of G (the number of clusters) and q (number of latent factors) in the algorithms to $(1, 2, 3)$ and $(1, 2, 3, 4)$ respectively.

Real data experiment II

- Both EM-GMCM and EM-GMM detect the correct number of clusters $G = 2$ with the CCC dependency structure and number of latent factors, $q = 3$.
- However, EM-GMCM outperforms other methods in classification accuracy.

Algorithm	EM-GMCM	EM-GMM	K-Means	Spectral	HEU-T
Accuracy	**72%**	59%	57%	54%	51%

EM-GMCM

	A	B
A	**814**	186
B	370	**630**

EM-GMM

	A	B
A	731	269
B	542	458

K-Means

	A	B
A	857	143
B	712	288

Spectral

	A	B
A	1000	0
B	993	7

HEU-T

	A	B
A	668	332
B	645	355

Table: Above: Classification accuracy of the algorithms tested on the Gamma Telescope Data. Below: Classification tables of the best models chosen by EM-GMCM and EM-GMM and of K-means, Spectral clustering and HEU-T. A: gamma signal, B: hadron showers (background). True labels: horizontal, Estimated labels: vertical.

Summary

- GMCM: models heterogeneous multimodal data where the dependence structure comes from a GMM
- GMCM based clustering outperforms popular approaches like GMM, K-means in our experiments
- Future work: high-dimensional, and categorical data

Research Challenges

- Better estimation algorithms
- Discrete Data
  - Sklar's Theorem: copula not unique
  - Hoff's extended rank likelihood approach
- High Dimensions

Copulas: higher dimensions

- A multivariate density may be decomposed into conditional densities:

$$f(x_1,\dots,x_n) = f_n(x_n).f(x_{n-1}|x_n)\dots f(x_1|x_2,\dots x_n)$$

- Each term in the above expression can be written as functions of bivariate copula densities:

$$f(x|v_1,\dots,v_d) = c_{xv_j|v_{-j}}(F(x|v_{-j}),F(v_j|v_{-j})).f(x|v_{-j}),$$

  where v_{-j} denotes the d-dimensional vector v excluding the j^{th} component.

- Example, $f(x_1|x_2,x_3) = c_{13|2}(F(x_1|x_2),F(x_3|x_2))f(x_1|x_2)$ and $f(x_2|x_3) = c_{23}(F(x_2),F(x_3)).f(x_2)$.

Copulas: higher dimensions

- The conditional distributions in the pair copula constructions, also called *h-functions*, are given by:

$$F(x|v) = \frac{\partial C_{x,v_j|v_{-j}}(F(x|v_{-j}), F(v_j|v_{-j}))}{\partial F(v_j|v_{-j})}.$$

- Analytic expressions for these h-functions have been derived for commonly used copulas. Thus all the densities in

$$f(x_1,\ldots,x_n) = f_n(x_n).f(x_{n-1}|x_n)\ldots f(x_1|x_2,\ldots x_n)$$

may be expressed in terms of univariate marginals and bivariate copulas.

- This forms the basis of pair copula constructions for multivariate distributions.

Vines: hierarchical collection of copulas

A D-vine has $n-1$ hierarchical trees and $n(n-1)/2$ bivariate copulas for n random variables. For example,

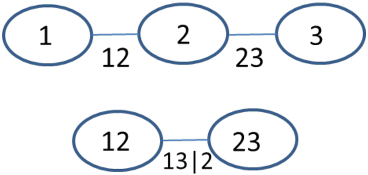


Figure: Three dimensional D Vine.

$$f(x_1,x_2,x_3) = f_1(x_1).f_2(x_2).f_3(x_3).c_{12}(F_1(x_1),F_2(x_2)).c_{23}(F_2(x_2),F_3(x_3)).\\ c_{13|2}(F(x_1|x_2),F(x_3|x_2))$$

Thank you!

References

- Joe, Harry. Dependence Modeling with Copulas. CRC Press, 2014.
- Kjerti Aas. Modelling the dependence structure of financial assets: A survey of four copulas. 2004.
- Christian Genest and Anne–Catherine Favre. Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. Journal of Hydrologic Engineering 2007.
- Thorsten Schmidt. Coping with Copulas. Copulas-From theory to application in finance (2007): 3-34.
- Gal Elidan. Copulas in Machine Learning. Copulae in Mathematical and Quantitative Finance. Springer Berlin Heidelberg, 2013. 39-60.
- Rey, Melanie, and Volker Roth. Copula mixture model for dependency-seeking clustering. ICML 2012.
- Poczos, Barnabas, Zoubin Ghahramani, and Jeff Schneider. Copula-based kernel dependency measures. ICML 2012.

References

- Sakyajit Bhattacharya and Vaibhav Rajan. Unsupervised Learning using Gaussian Mixture Copula Model. In 21st International Conference on Computational Statistics (COMPSTAT 2014), Geneva, Switzerland.
- A. Tewari, M. J. Giering, and A. Raghunathan. Parametric characterization of multimodal distributions with non-gaussian modes. In Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops, pages 286–292, 2011.
- K. Bache and M. Lichman. UCI machine learning repository, 2013.
- G. McLachan and K. Basford. Mixture Models: Inference and Applications to Clustering. Marcel Dekker Inc., 1988.
- P. D. McNicholas and T. Murphy. Parsimonious Gaussian mixture models. Statistics and Computing, 18(3):285–296, 2008.