Presentation:

GAODE and HAODE:

Two Proposals based on AODE to Deal with Continuous Variables

International Conference on Machine Learning 2009
on 15/06/2009

M. Julia Flores, José A. Gámez, Ana M. Martínez and José M. Puerta
Computing Systems Department
Albacete - UCLM - Spain

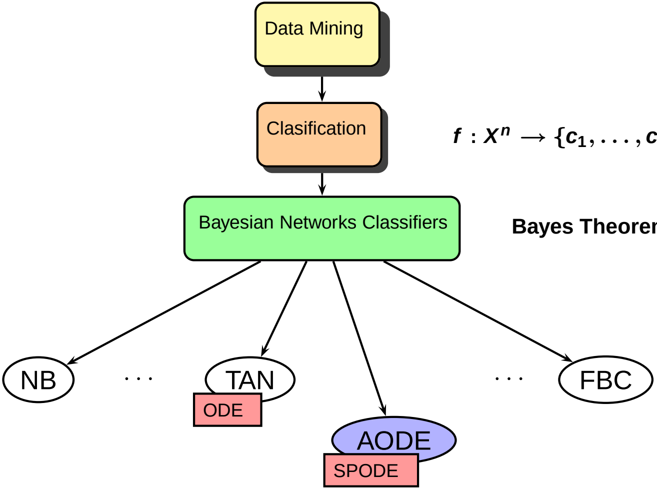
Outline

1. **Motivation**
2. **Conditional Gaussian Networks**
3. **New proposals to deal with numeric attributes in AODE**
   - Gaussian AODE (GAODE) classifier
   - Hybrid AODE (HAODE) classifier
4. **Experimental Methodology and Results**
   - Datasets with only continuous attributes
   - Hybrid Datasets
5. **Conclusions and Future Work**

Outline

1. **Motivation**
2. **Conditional Gaussian Networks**
3. **New proposals to deal with numeric attributes in AODE**
   - Gaussian AODE (GAODE) classifier
   - Hybrid AODE (HAODE) classifier
4. **Experimental Methodology and Results**
   - Datasets with only continuous attributes
   - Hybrid Datasets
5. **Conclusions and Future Work**

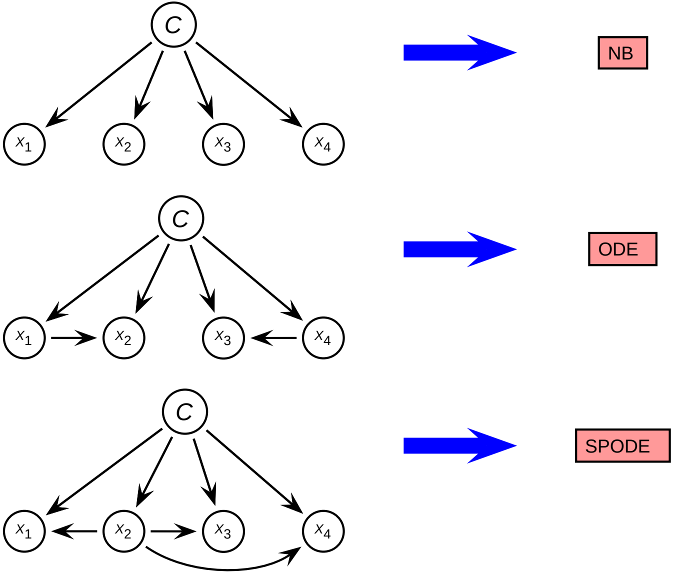
Framework



$$f : X^n \longrightarrow \{c_1, \ldots, c_k\}$$

Bayes Theorem

Different network structures



Naive Bayes classifier

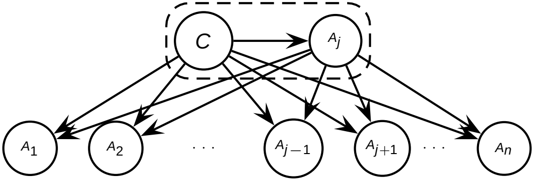
- The attributes are conditionally independent given the class value $I(A_i, A_j | C)$.

$$c_{MAP} = argmax_{c \in \Omega_C} \; p(c) \prod_{i=1}^{n} p(a_i | c)$$

- **Time complexity**:
  - **Training**: $\mathcal{O}(tn)$
  - **Clasification**: $\mathcal{O}(kn)$
- Problems:
  - ✕ : It does not work properly in certain datasets.
  - ✕ : Dependences between attributes reduce, unavoidable, the prediction capability of NB.
  - ✕ : Not only interesting to be right in the classification in certain applications.

AODE classifier I

- **AODE** is significantly better in terms of error reduction compared to the rest of semi-naive techniques.



- **MAP hypothesis:**

$$argmax_{c \in \Omega_C} \left(\sum_{j=1, N(a_j) > m}^{n} p(c, a_j) \prod_{i=1, i \neq j}^{n} p(a_i | c, a_j) \right)$$

AODE classifier II

- **Time complexity**:
  - **Training**: $\mathcal{O}(tn^2)$
  - **Classification**: $\mathcal{O}(kn^2)$
- Drawbacks:
  - × : Quadratical order time in classification.
  - × : High demand of RAM memory.
  - × : **Only discrete variables**.

ECSQARU 09

- Attemps to improve AODE's **accuracy**
  - **WAODE**: Model weighting with $IM(C, A_j)$.

Motivation

- A BN assume **all** the variables are **discrete**.
- **Large amount of methods** developed to solve problems with **discrete variables**.
- It is common the **coexistence** of **discrete** and **continuous** variables in the same problem.
- Direct solution $\longrightarrow$ **discretization**.
  - ✗ : Unavoidable lost of precision.
  - ✗ : Which discretization method should we choose?.
- Some alternative solutions:
  1. **Conditional Gaussian Networks**.
  2. **Kernel-based distributions**.
  3. **Mixtures of truncated exponentials**.

Outline

1. Motivation
2. **Conditional Gaussian Networks**
3. New proposals to deal with numeric attributes in AODE
   - Gaussian AODE (GAODE) classifier
   - Hybrid AODE (HAODE) classifier
4. Experimental Methodology and Results
   - Datasets with only continuous attributes
   - Hybrid Datasets
5. Conclusions and Future Work

Conditional Gaussian Networks

- Every **continuous variable** X is modeled with a **conditional Gaussian distribution**.

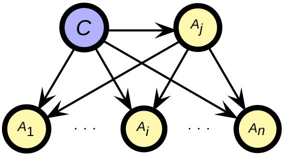
$$f(X|\mathbf{Y}=y,\mathbf{Z}=z;\Theta)=\mathcal{N}(x:\mu_X(y)+\sum_{j=1}^{s} b_{XZ_j}(y)(z_j-\mu_{Z_j}(y)),\sigma^2_{X|\mathbf{Z}}(y))$$

- $b_{XZ_j}(y)$, **regression term** that measures the strength of the connection between X and every continuous parent.

- $\sigma^2_{X|\mathbf{Z}}(y)$ is the **conditional variance** of X over its continuous parents.

Factorization of the joint density function in a SPODE network

Structure



Local densities

$\Theta_C = (CPT)$	$f_C \rightsquigarrow P(C)$
$\Theta_j = (\mu_j(c), -, \sigma_j(c))$	$f_{A_j \mid C=c} \rightsquigarrow \mathcal{N}(\mu_j(c), \sigma_j(c))$
$\Theta_1 = (\mu_1(c), b_1(c), \sigma_{1\mid j}(c))$	$f_{A_1 \mid C=c, A_j=a_j} \rightsquigarrow \mathcal{N}(\mu_1(c) + b_{1j}(c)(a_j - \mu_j(c))), \sigma_{1\mid j}(c))$
$\Theta_i = (\mu_i(c), b_i(c), \sigma_{i\mid j}(c))$	$f_{A_i \mid C=c, A_j=a_j} \rightsquigarrow \mathcal{N}(\mu_i(c) + b_{ij}(c)(a_j - \mu_j(c))), \sigma_{i\mid j}(c))$
$\Theta_n = (\mu_n(c), b_n(c), \sigma_{n\mid j}(c))$	$f_{A_n \mid C=c, A_j=a_j} \rightsquigarrow \mathcal{N}(\mu_n(c) + b_{nj}(c)(a_j - \mu_j(c))), \sigma_{n\mid j}(c))$

Factorization of the joint density function

$$f(c, a_j, a_1, \ldots, a_i, \ldots, a_n) = p(c) \frac{1}{\sqrt{2\pi}\sigma_j(c)} e^{-\frac{1}{2}\left(\frac{a_j - \mu_j(c)}{\sigma_j(c)}\right)^2} \cdot$$

$$\cdot \frac{1}{\sqrt{2\pi}\sigma_{i\mid j}(c)} e^{-\frac{1}{2}\left(\frac{a_i - (\mu_i(c) + b_{ij}(a_j - \mu_j(c)))}{\sigma_{i\mid j}(c)}\right)^2} \cdots \frac{1}{\sqrt{2\pi}\sigma_{n\mid j}(c)} e^{-\frac{1}{2}\left(\frac{a_n - (\mu_n(c) + b_{nj}(a_j - \mu_j(c)))}{\sigma_{n\mid j}(c)}\right)^2}$$

Outline

1. Motivation
2. Conditional Gaussian Networks
3. **New proposals to deal with numeric attributes in AODE**
   - Gaussian AODE (GAODE) classifier
   - Hybrid AODE (HAODE) classifier
4. Experimental Methodology and Results
   - Datasets with only continuous attributes
   - Hybrid Datasets
5. Conclusions and Future Work

Outline

1. Motivation

2. Conditional Gaussian Networks

3. **New proposals to deal with numeric attributes in AODE**
   - Gaussian AODE (GAODE) classifier
   - Hybrid AODE (HAODE) classifier

4. Experimental Methodology and Results
   - Datasets with only continuous attributes
   - Hybrid Datasets

5. Conclusions and Future Work

GAODE classifier

- Application of **CGN** to the children's distributions:
  - The density function just has a **discrete parent node** (***C***) and a **continous parent node**, which is the superparent on each model (***A_j***).

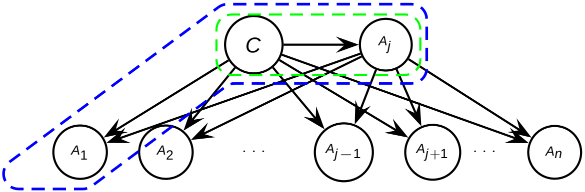
$$f(A_i = a_i | C = c, A_j = a_j) = \mathcal{N}(a_i : \mu_i(c) + b_{ij}(c)(a_j - \mu_j(c)), \sigma^2_{i|j}(c))$$

- **MAP hypothesis:**

$$argmax_c \left(\sum_{j=1}^{n} \mathcal{N}(a_j : \mu_j(c), \sigma_j^2(c)) p(c) \prod_{i=1 \wedge i \neq j}^{n} \mathcal{N}(a_i : \mu_i(c) + b_{ij}(c)(a_j - \mu_j(c)), \sigma^2_{i|j}(c)) \right)$$

GAODE classifier

- Univariate Gaussian Distribution -



- Conditional Gaussian Distribution -

GAODE classifier

- **Time complexity**:
  - The same as AODE (incremental computation of parameters).
- **Space complexity**:
  - **Training & Classification**: $\mathcal{O}(kn^2)$ (independent from v).

√ Probabilities estimated can be more reliable comparing to the multinomial version as they are modeled from more samples, especially with large CPTs.

× : Not possible to define the corresponding probability function for a discrete variable conditioned to a numeric attribute. **Restricted to Numerical datasets**.

Outline

1. Motivation

2. Conditional Gaussian Networks

3. **New proposals to deal with numeric attributes in AODE**
   - Gaussian AODE (GAODE) classifier
   - Hybrid AODE (HAODE) classifier

4. Experimental Methodology and Results
   - Datasets with only continuous attributes
   - Hybrid Datasets

5. Conclusions and Future Work

HAODE classifier

- **Discrete** superparents ($\boldsymbol{A_j}$) in its corresponding models.
- **MAP hypothesis:**.

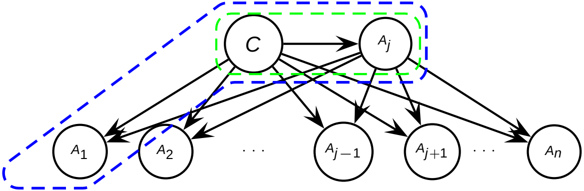
$$argmax_c \left(\sum_{j=1, N(a_j) > m}^{n} p(a_j, c) \prod_{i=1 \wedge i \neq j}^{n} \mathcal{N}(a_i : \mu_i(c, a_j), \sigma_i^2(c, a_j)) \right)$$

- **Time complexity**:
  - The same as AODE (incremental computation of parameters).
- **Space complexity**:
  - **Training & Classification**: The same as AODE $\boldsymbol{\mathcal{O}(k(nv)^2)}$ in the worst of the cases.

√ Able to deal with **hybrid datasets** too.

HAODE classifier

- Multinomial distribution -



- Univariate Gaussian distribution -

Outline

1. Motivation
2. Conditional Gaussian Networks
3. New proposals to deal with numeric attributes in AODE
   - Gaussian AODE (GAODE) classifier
   - Hybrid AODE (HAODE) classifier
4. **Experimental Methodology and Results**
   - Datasets with only continuous attributes
   - Hybrid Datasets
5. Conclusions and Future Work

Outline

1. Motivation

2. Conditional Gaussian Networks

3. New proposals to deal with numeric attributes in AODE
   - Gaussian AODE (GAODE) classifier
   - Hybrid AODE (HAODE) classifier

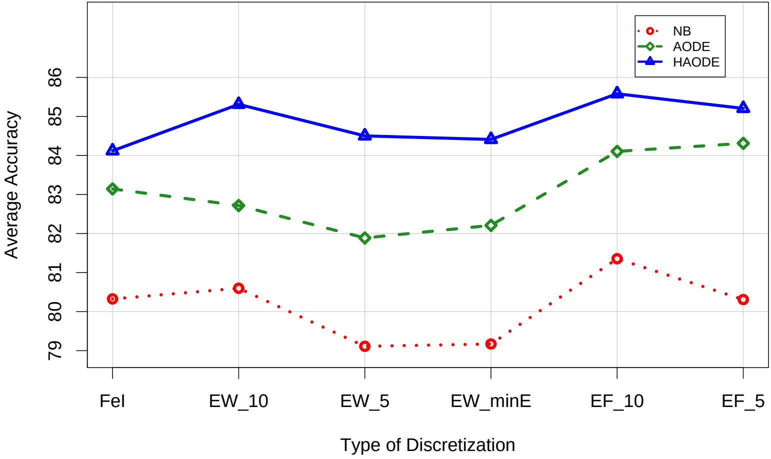
4. **Experimental Methodology and Results**
   - **Datasets with only continuous attributes**
   - Hybrid Datasets

5. Conclusions and Future Work

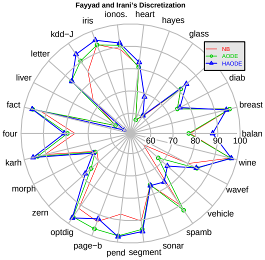
Experimental framework

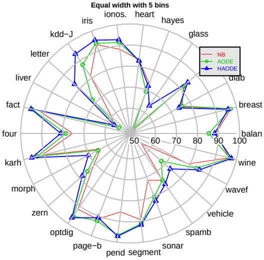
- Experiments over the **26 numeric datasets** for clasification, downloaded from the Weka home page and original from the UCI repository.
- 5x2*cv* for the evaluation process.
- 5x2 cv F Test with a 95 %.
- Supervised discretization for NB, AODE and HAODE's superparents.
  - Fayyad and Irani's MDL method.
  - Further experiments have been performed with different discretization methods, and the results obtained follow the same tendency.

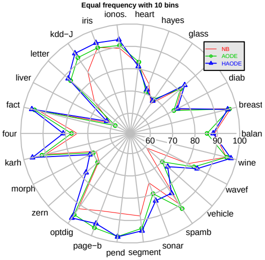
Comparison between different discretization methods

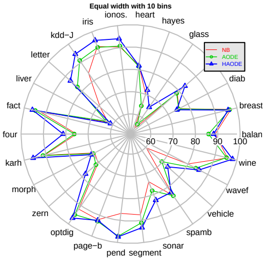


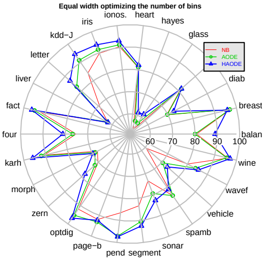
Comparison between different discretization methods

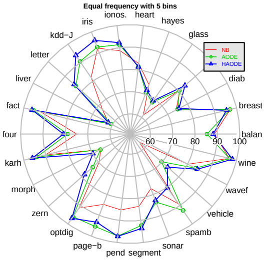












Datasets with only continuous attributes

Id	NB-G	NB	AODE	**GAODE**	**HAODE**
balance-scale	•88, 864	77, 632	76, 992	•89, 088	•87, 68
breast-w	•96, 0801	•97, 1102	•96, 6237	•95, 9662	•95, 0787
diabetes	•74, 974	•74, 6875	•74, 5573	•74, 7917	•75, 9115
ecoli	•83, 9881	•80, 7738	•81, 0119	•84, 5238	•84, 3452
glass	49, 7196	•60	•60, 7477	•52, 8037	•60, 6542
hayes-roth	•65, 375	•57, 5	•57, 5	•65, 625	•68, 5
heart-statlog	•83, 4815	•81, 2593	•80, 8148	•83, 7778	•83, 037
ionosphere	•82, 963	•88, 8889	•90, 7123	•92, 0228	•91, 7379
iris	•95, 0667	•93, 4667	•93, 3333	•97, 4667	•95, 6
kdd-JapanV	85, 7444	84, 5758	90, 3885	91, 8442	•93, 9966
letter	64, 06	73, 296	•86, 292	71, 235	•86, 138
liver-disorders	•54, 2609	•58, 6087	•58, 6087	•57, 3333	•54, 2029
mfeat-factors	92, 29	92, 36	•96, 08	•95, 94	•96, 31
mfeat-fourier	75, 7	75, 87	79, 25	•79, 39	•80, 69
mfeat-karh	93, 16	90, 48	•93, 83	•96, 15	•95, 92
mfeat-morph	•69, 32	68, 03	68, 9	•70, 79	•69, 95
mfeat-zernike	72, 99	70, 21	74, 63	•77, 42	•78, 1
optdigits	91, 1317	91, 7544	•96, 3167	93, 637	•96, 9181
page-blocks	•87, 7142	93, 1336	•96, 6307	•90, 9446	•91, 8144
pendigits	85, 7041	87, 3362	•97, 1161	94, 2085	•97, 5182
segment	80, 6753	90, 4416	•94, 1732	86, 6667	•95, 1602
sonar	67, 5	•75, 6731	•75, 5769	•71, 4423	•75, 9615
spambase	79, 5131	89, 8544	•92, 7277	79, 8566	77, 3658
vehicle	43, 1678	58, 6052	67, 4704	•68, 5106	•72, 9787
waveform-5000	80	79, 968	•84, 508	•84, 46	•84, 22
wine	97, 4157	96, 9663	•96, 9663	•98, 427	97, 4157
Av	**78, 4842**	**80, 3262**	**83, 1445**	**82, 4739**	**84, 1233**

Datasets with only continuous attributes

Ftest Wilcoxon	NB-G	NB	AODE	**GAODE**
NB	7-16-3 NO			
AODE	11-14-1 AODE	14-12-0 AODE		
GAODE	12-14-0 GAODE	12-12-2 GAODE	5-16-5 NO	
HAODE	13-13-0 HAODE	13-12-1 HAODE	6-19-1 HAODE	6-18-2 HAODE

- √ Both of our classifiers are significantly better than NB.
- √ **GAODE** obtains compatitive results comparing to AODE.
- √ **HAODE** offers an even higher advantage, HAODE **significantly improves AODE in numeric datasets**.

Friedman and Nemenyi tests

- **Friedman** statistic (chi-square with 4 df: 27,12): $1{,}877e-05$.
- Iman and Davenport statistic (F-distribution with 4 and 100 df: 8,82): $3{,}852e-06$.
- **Nemenyi** tests: only rejected the null hypothesis in favor of GAODE and HAODE over NB-G and NB.

Table: Adjusted *p*-values

i	hypothesis	unadjusted *p*	p_{Neme}	p_{Holm}
1	"NB-G"vs ."HAODE"	2.555080653283355E-5	2.555080653283355E-4	2.555080653283355E-4
2	"NB"vs ."HAODE"	2.299235858400624E-4	0.002299235858400624	0.0020693122725605616
3	"NB-G"vs ."GAODE"	3.8219140804744933E-4	0.0038219140804744934	0.0030575312643795947
4	"NB"vs ."GAODE"	0.002479351316213918	0.02479351316213918	0.017355459213497428
5	"NB-G"vs ..AODE"	0.01405860532373715	0.1405860532373715	0.0843516319424229
6	"NB"vs ..AODE"	0.053665391826160654	0.5366539182616066	0.2683269591308033
7	.AODE"vs ."HAODE"	0.07941062599894239	0.7941062599894239	0.31764250399576954
8	.AODE"vs ."GAODE"	0.27293765570660644	2.7293765570660646	0.8188129671198193
9	"GAODE"vs ."HAODE"	0.5106708223861578	5.1067082238615775	1.0213416447723156
10	"NB-G"vs ."NB"	0.598725069652846	5.98725069652846	1.0213416447723156

Outline

1. Motivation
2. Conditional Gaussian Networks
3. New proposals to deal with numeric attributes in AODE
   - Gaussian AODE (GAODE) classifier
   - Hybrid AODE (HAODE) classifier
4. **Experimental Methodology and Results**
   - Datasets with only continuous attributes
   - Hybrid Datasets
5. Conclusions and Future Work

Hybrid Datasets

- As HAODE is able to deal with **all kind of datasets**.
- Experiments with the 16 hybrid datasets of the **standard group of 36 datasets** from the UCI repository.
- Same Experimental framework.

Id	NB	AODE	HAODE	%*M*
zoo	●90, 495	●91, 6832	●94, 2574	0
lymph	●81, 0811	●80, 8108	●82, 5676	0
vowel	50, 6667	61, 0505	●78, 4444	0
credit-g	●74, 16	●74, 44	●75, 32	0
labor	●88, 4211	●87, 7193	●88, 0702	0
anneal	●95, 1448	●96, 7483	●92, 784	0
heart-c	●83, 3003	●83, 3003	●83, 7624	0, 17
credig-a	●86, 029	●86, 2609	78, 8696	5
hepatitis	●82, 3226	●83, 0968	●84, 3871	5, 39
hypothyroid	●97, 7253	●98, 0011	95, 6416	5, 4
sick	97, 0891	●97, 2057	94, 5652	5, 4
autos	●58, 7317	●64, 1951	●57, 561	11, 06
colic.ORIG	●69, 6196	●69, 7826	60, 8696	18, 7
heart-h	●83, 8776	●83, 9456	●83, 4014	19
colic	●79, 3478	●81, 087	●78, 8043	22, 77
anneal.ORIG	●93, 1403	●93, 9866	88, 7751	63, 32
Av	●81, 947	●83, 3321	●82, 3801	

What happen in hybrid datasets?

- Results summary:
  - **HAODE-ties-AODE**: 1 – 10 – 5.
  - *Wilcoxon test*: No statistical diference.
- **Why?**
  - No significant pattern found when the percentage of numerical variables with respect to discrete ones was analyzed.
  - Apparent tendency of HAODE to **punnish datasets with missing values**. Statistical difference (Wilcoxon) when only datasets with missing values are considered.
  - Results summary after applying an unsupervised filter to replace missing values: 2 – 12 – 2.

Outline

1. Motivation
2. Conditional Gaussian Networks
3. New proposals to deal with numeric attributes in AODE
   - Gaussian AODE (GAODE) classifier
   - Hybrid AODE (HAODE) classifier
4. Experimental Methodology and Results
   - Datasets with only continuous attributes
   - Hybrid Datasets
5. **Conclusions and Future Work**

Conclusions and Future Work

- In this paper we propose two alternatives to AODE to deal with continuous attributes.
- **GAODE**: applies CGNs.
  - √ **Competitive results** comparing to AODE (5 – 16 – 5).
  - √ **Reduction** in the **space complexity**.
  - √ Maintains AODE's time complexity.
  - √ Can entail a more reliable computation of statistics, only class conditioned.
  - × : Restricted to continuous datasets.
- **HAODE**: discretizes the superparents.
  - √ **Significantly better** than AODE in continuous datasets (6 – 19 – 1).
  - √ Able to deal with **all kind of datasets**.
  - × : Clear preference for datasets with continuous attributes and absence of missing data.
- As future line of work: application of Kernel-based distributions, **Mixtures of Truncated Exponentials** (MTEs) and other approaches.

Thank you