



Presentación:

Estudio y Comparativa de Diferentes Discretizaciones en Clasificadores Bayesianos

*Conferencia de la Asociación Española para la Inteligencia
Artificial 2009*

11/11/2009

M. Julia Flores, José A. Gámez, Ana M. Martínez and
José M. Puerta

Departamento de Sistemas Informáticos
Albacete - UCLM - Spain



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

Outline

1 Motivación

2 Clasificadores basados en Redes Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE (HAODE)

3 Métodos de discretización

4 Metodología experimental y resultados

Estudio de la tasa de acierto

Estudio del sesgo y la varianza

5 Conclusiones y trabajo futuro



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

Outline

1 Motivación

2 Clasificadores basados en Redes Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE (HAODE)

3 Métodos de discretización

4 Metodología experimental y resultados

Estudio de la tasa de acierto

Estudio del sesgo y la varianza

5 Conclusiones y trabajo futuro



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

- La **discretización** es probablemente una de las técnicas de **preprocesamiento más utilizadas** en aprendizaje automático.
- La distribución real de los datos es sustituida por una mixtura de uniformes.
- **Motivos:**
 - Métodos **restringidos** a variables **discretas**.
 - **Disminución** en el **tiempo** de ejecución de un algoritmo.
 - **Reducción** del **ruido** presente en los datos originales.
- Existen distintas técnicas de discretización.
- *Poco habitual justificar el método de discretización elegido.*



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

- **Análisis empírico** de este problema:
 - Subconjunto de clasificadores basados en RBs: **Naive Bayes**, **AODE** y **HAODE**.
 - **Técnicas de discretización supervisadas** (Fayyad e Irani) y **no supervisadas** (igual frecuencia y anchura).
- **Objetivo:** analizar un conjunto de discretizaciones y comprobar si el resultado obtenido por los clasificadores es **sensible** al método de discretización utilizado.
- **Resultado:** existen alteraciones en cuanto al grado de exactitud dependiendo de la discretización, pero **la tendencia es la misma**, esto es, cambia el tanto por ciento de acierto, pero no el ranking entre clasificadores.



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

Outline

1 Motivación

2 Clasificadores basados en Redes Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE (HAODE)

3 Métodos de discretización

4 Metodología experimental y resultados

Estudio de la tasa de acierto

Estudio del sesgo y la varianza

5 Conclusiones y trabajo futuro



Clasificación

$$f : X^n \rightarrow \{c_1, \dots, c_k\}$$

Ana M. Martínez



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

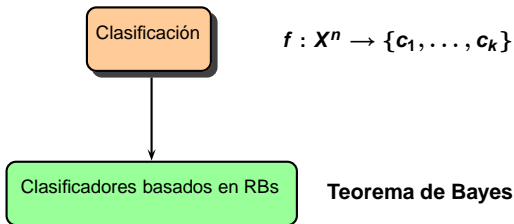
Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias



Motivación

Clasificadores basados en Redes Bayesianas

- Naive Bayes
- AODE
- Clasificador Hybrid AODE
(HAODE)

Métodos de discretización

Metodología experimental y resultados

- Estudio de la tasa de
acierto
- Estudio del sesgo y la
varianza

Conclusiones y trabajo futuro

Referencias



Motivación

Clasificadores basados en Redes Bayesianas

Naive Bayes
AODE
Clasificador Hybrid AODE
(HAODE)

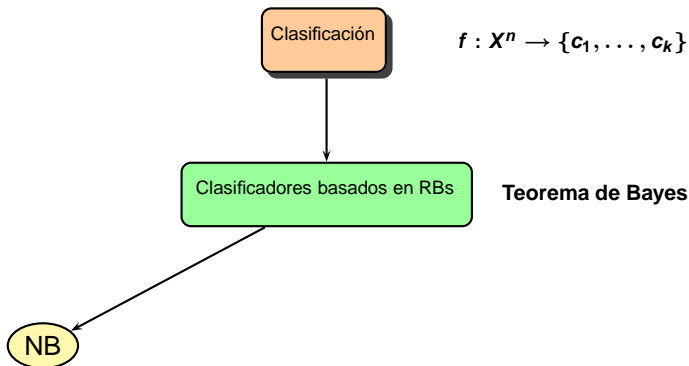
Métodos de
discretización

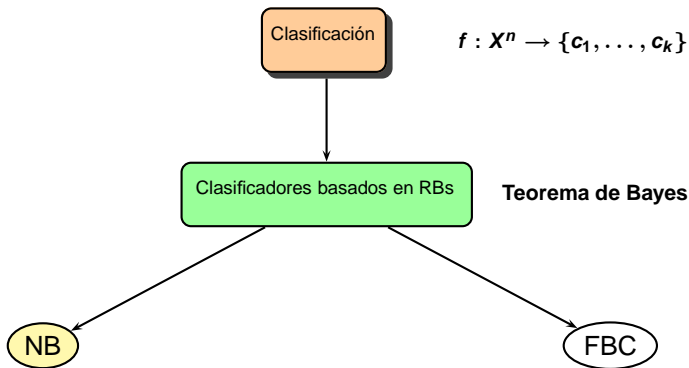
Metodología
experimental y
resultados

Estudio de la tasa de
acierto
Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias







Motivación

Clasificadores basados en Redes Bayesianas

- Naive Bayes
- AODE
- Clasificador Hybrid AODE (HAODE)

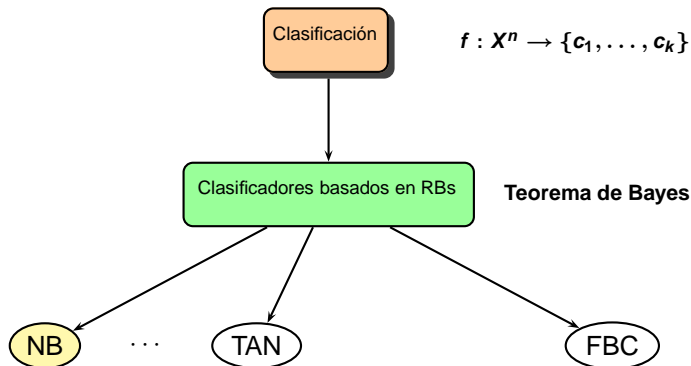
Métodos de discretización

Metodología experimental y resultados

- Estudio de la tasa de acierto
- Estudio del sesgo y la varianza

Conclusiones y trabajo futuro

Referencias





Motivación

Clasificadores basados en Redes Bayesianas

Naive Bayes
AODE
Clasificador Hybrid AODE
(HAODE)

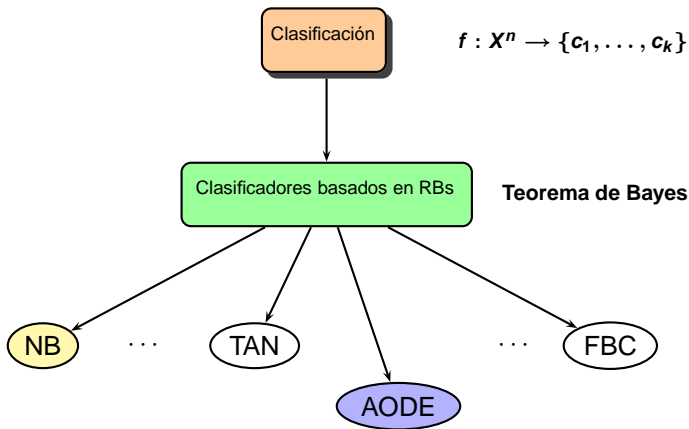
Métodos de discretización

Metodología experimental y resultados

Estudio de la tasa de
acierto
Estudio del sesgo y la
varianza

Conclusiones y trabajo futuro

Referencias





Motivación

Clasificadores basados en Redes Bayesianas

- Naive Bayes
- AODE
- Clasificador Hybrid AODE (HAODE)

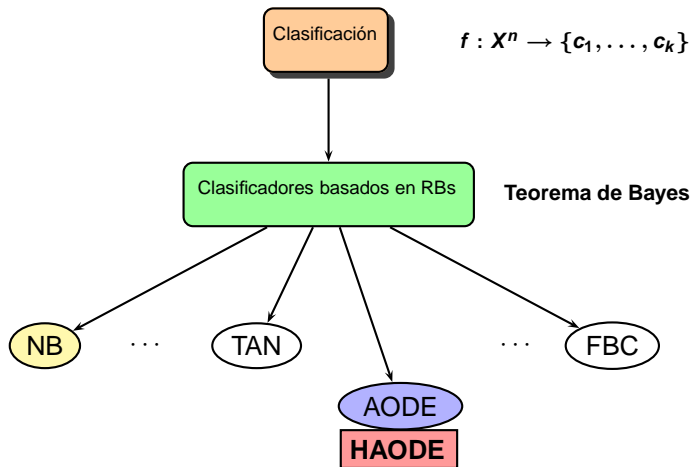
Métodos de discretización

Metodología experimental y resultados

- Estudio de la tasa de acierto
- Estudio del sesgo y la varianza

Conclusiones y trabajo futuro

Referencias



Outline

1 Motivación

2 Clasificadores basados en Redes Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE (HAODE)

3 Métodos de discretización

4 Metodología experimental y resultados

Estudio de la tasa de acierto

Estudio del sesgo y la varianza

5 Conclusiones y trabajo futuro



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

Clasificador Naive Bayes

- Los atributos son condicionalmente independientes conocido el valor de la clase $I(A_i, A_j|C)$.

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

- **Complejidad temporal:**

- **Entrenamiento:** $\mathcal{O}(tn)$
- **Clasificación:** $\mathcal{O}(kn)$

- **Problemas:**

- ✗ : Rendimiento pobre en ciertas bases de datos.
- ✗ : Las dependencias entre atributos reducen, inevitablemente, el poder de predicción de NB.
- ✗ : No sólo interesa conocer la clase predicha sino la certeza con la que se realiza la misma.



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

Outline

1 Motivación

2 Clasificadores basados en Redes Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE (HAODE)

3 Métodos de discretización

4 Metodología experimental y resultados

Estudio de la tasa de acierto

Estudio del sesgo y la varianza

5 Conclusiones y trabajo futuro



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

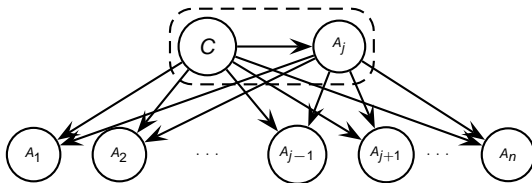
Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias



- **AODE** es significativamente mejor en términos de reducción del error comparado con el resto de técnicas semi-naive [ZW05].



- **Hipótesis MAP:**

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

Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

- **Complejidad temporal:**
 - **Entrenamiento:** $\mathcal{O}(tn^2)$
 - **Clasificación:** $\mathcal{O}(kn^2)$



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

- **Complejidad temporal:**
 - **Entrenamiento:** $\mathcal{O}(tn^2)$
 - **Clasificación:** $\mathcal{O}(kn^2)$

- **Desventajas:**



- **Complejidad temporal:**
 - **Entrenamiento:** $\mathcal{O}(tn^2)$
 - **Clasificación:** $\mathcal{O}(kn^2)$

- **Desventajas:**
 - ✗ : Orden cuadrático en la clasificación.
 - ✗ : Alta demanda de memoria RAM.



- **Complejidad temporal:**
 - **Entrenamiento:** $\mathcal{O}(tn^2)$
 - **Clasificación:** $\mathcal{O}(kn^2)$

- **Desventajas:**
 - ✗ : Orden cuadrático en la clasificación.
 - ✗ : Alta demanda de memoria RAM.
 - ✗ : **Sólo variables discretas.**



- **Complejidad temporal:**
 - Entrenamiento: $\mathcal{O}(tn^2)$
 - Clasificación: $\mathcal{O}(kn^2)$
- Desventajas:
 - ✗ : Orden cuadrático en la clasificación.
 - ✗ : Alta demanda de memoria RAM.
 - ✗ : **Sólo variables discretas.**

HODE
ECSQARU 09

Ana M. Martínez



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

- **Complejidad temporal:**
 - **Entrenamiento:** $\mathcal{O}(tn^2)$
 - **Clasificación:** $\mathcal{O}(kn^2)$
- **Desventajas:**
 - ✗ : Orden cuadrático en la clasificación.
 - ✗ : Alta demanda de memoria RAM.
 - ✗ : **Sólo variables discretas.**

HODE
ECSQARU 09

GAODE/HAODE
ICML 09

Ana M. Martínez



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

Outline

1 Motivación

2 Clasificadores basados en Redes Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE (HAODE)

3 Métodos de discretización

4 Metodología experimental y resultados

Estudio de la tasa de acierto

Estudio del sesgo y la varianza

5 Conclusiones y trabajo futuro



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

- Superpadres **discretos** (\mathbf{A}_j) en cada modelo.
- **Hipótesis MAP:**

$$\operatorname{argmax}_c \left(\sum_{j=1, N(\mathbf{a}_j) > m}^n p(\mathbf{a}_j, c) \prod_{i=1 \wedge i \neq j}^n \mathcal{N}(\mathbf{a}_i : \mu_i(c, \mathbf{a}_j), \sigma_i^2(c, \mathbf{a}_j)) \right)$$

- **Complejidad temporal:**
 - **La misma que AODE** (cálculo incremental de parámetros).
 - **Complejidad espacial:**
 - **Entrenamiento & Clasificación:** la misma que AODE $\mathcal{O}(k(nv)^2)$ en el peor de los casos.
- ✓ Capaz de manejar **bases de datos híbridas** también.



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes
AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

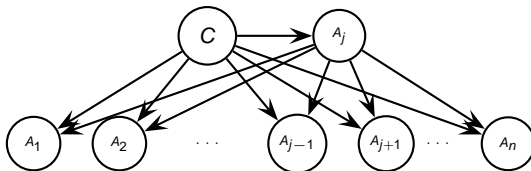
Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

Clasificador HAODE



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

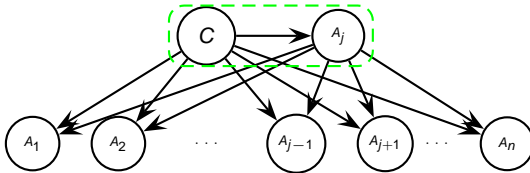
Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

Clasificador HAODE



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

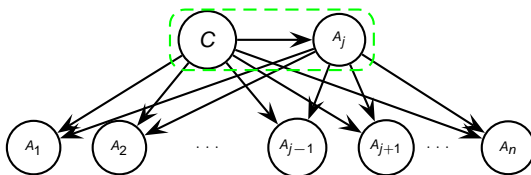
Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias



- Distribución Multinomial -



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes
AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

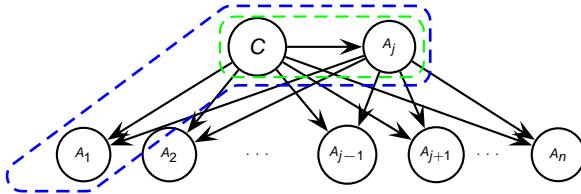
Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

- Distribución Multinomial -



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

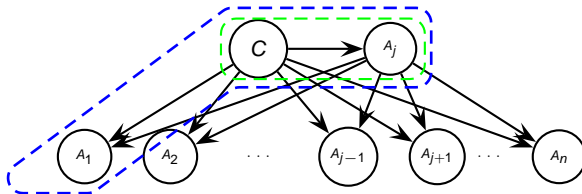
Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias



- Distribución Multinomial -



- Distribución Gaussiana Univariada -

Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes
AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

Outline

1 Motivación

2 Clasificadores basados en Redes Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE (HAODE)

3 Métodos de discretización

4 Metodología experimental y resultados

Estudio de la tasa de acierto

Estudio del sesgo y la varianza

5 Conclusiones y trabajo futuro



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

Métodos de discretización evaluados

- **Discretización por división de intervalos con igual anchura** (no supervisada)
 - División en un número de intervalos b de **igual anchura**.
 - Habitual fijar este valor en **5 ó 10 intervalos**.
 - **Valor más adecuado** de b de forma filter **minimizando la entropía** de la partición.
- **Discretización por división de intervalos con igual frecuencia** (no supervisada)
 - División en b intervalos con t/b instancias.
 - Complejidad temporal de ambas técnicas: $\mathcal{O}(t \log t)$.
- **Discretización basada en mínima entropía de Fayyad e Irani** (supervisada)
 - Discretización binaria con el punto de corte que **minimice la entropía** entre los candidatos.
 - Repetición recursiva aplicándose el **criterio MDL** para decidir cuando finalizar.
 - Complejidad temporal: $\mathcal{O}(kt \log t)$.



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

Outline

- 1 Motivación
- 2 Clasificadores basados en Redes Bayesianas
 - Naive Bayes
 - AODE
 - Clasificador Hybrid AODE (HAODE)
- 3 Métodos de discretización
- 4 Metodología experimental y resultados
 - Estudio de la tasa de acierto
 - Estudio del sesgo y la varianza
- 5 Conclusiones y trabajo futuro



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

Marco experimental I

- Experimentos sobre **26 bases de datos numéricas** (página web de Weka y repositorio UCI).

Tabla: Características principales de las bases de datos: número de variables predictoras (n), número de clases (k), y número de instancias (t).

Id	Base de datos	n	k	t	Id	Base de datos	n	k	t
1	balance-scale	4	3	625	14	mfeat-fourier	76	10	2000
2	breast-w	9	2	699	15	mfeat-karh	64	10	2000
3	diabetes	8	2	768	16	mfeat-morph	6	10	2000
4	ecoli	7	8	336	17	mfeat-zernike	47	10	2000
5	glass	9	7	214	18	optdigits	64	9	5620
6	hayes-roth	4	4	160	19	page-blocks	10	5	5473
7	heart-statlog	13	2	270	20	pendigits	16	9	10992
8	ionosphere	34	2	351	21	segment	19	7	2310
9	iris	4	3	150	22	sonar	60	2	208
10	kdd-JapanV	14	9	9961	23	spambase	57	2	4601
11	letter	16	26	20000	24	vehicle	18	4	946
12	liver-disorders	6	2	345	25	waveform-5000	40	3	5000
13	mfeat-factors	216	10	2000	26	wine	13	3	178



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

- Validación cruzada 5x2.
- **6 tipos de discretización** de uso común:
 - 1 División en intervalos de igual anchura con 5 intervalos (DIA_5).
 - 2 División en intervalos de igual anchura con 10 intervalos (DIA_10).
 - 3 Optimización del número de intervalos mediante la minimización de la entropía (DIA_mE).
 - 4 División en intervalos de igual frecuencia con 5 intervalos (DIF_5).
 - 5 División en intervalos de igual frecuencia con 10 intervalos (DIF_10).
 - 6 Método heurístico de mínima entropía de Fayyad e Irani (Fel).
- **Filtros de discretización** incluidos en **Weka**.
- Clasificadores NB, AODE y HAODE.



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

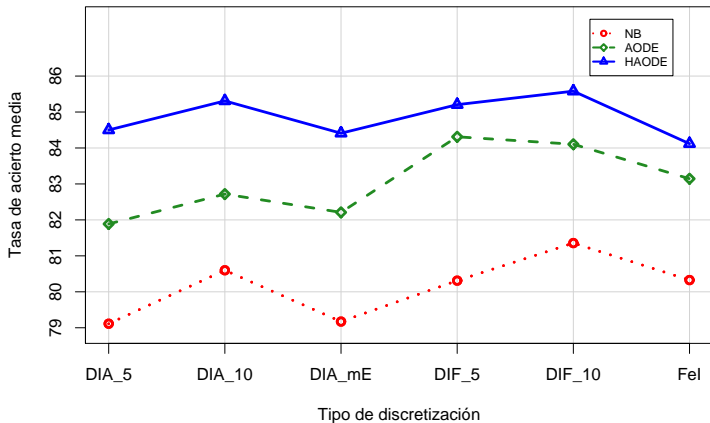
Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

Estudio de la tasa de acierto

- Porcentaje de aciertos en el conjunto de datos de test.



Motivación

Clasificadores basados en Redes Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE (HAODE)

Métodos de discretización

Metodología experimental y resultados

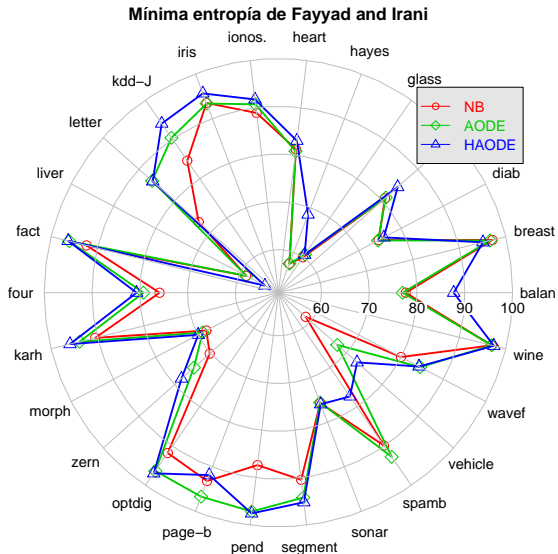
Estudio de la tasa de acierto

Estudio del sesgo y la varianza

Conclusiones y trabajo futuro

Referencias

Estudio de la tasa de acierto



Motivación

Clasificadores basados en Redes Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE (HAODE)

Métodos de discretización

Metodología experimental y resultados

Estudio de la tasa de acierto

Estudio del sesgo y la varianza

Conclusiones y trabajo futuro

Referencias

Estudio de la tasa de acierto

Ana M. Martínez



Motivación

Clasificadores basados en Redes Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE (HAODE)

Métodos de discretización

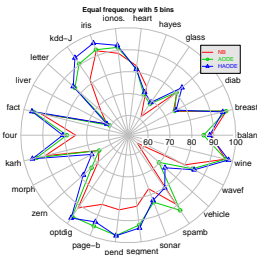
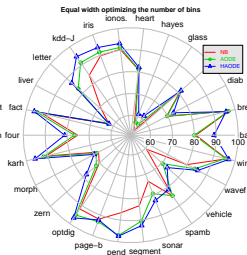
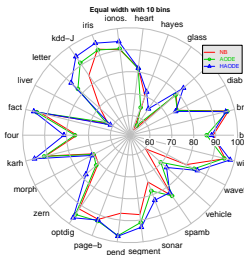
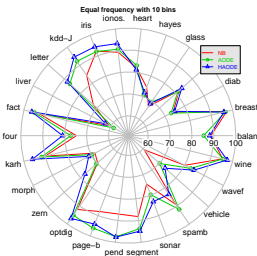
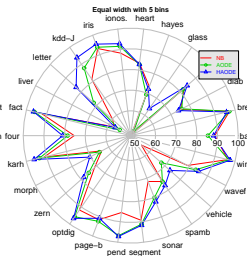
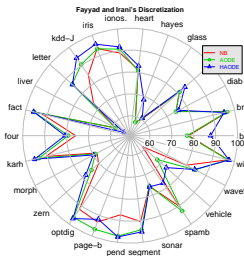
Metodología experimental y resultados

Estudio de la tasa de acierto

Estudio del sesgo y la varianza

Conclusiones y trabajo futuro

Referencias



Estudio de las diferentes discretizaciones:

- **Tests de Friedman** para la comparación múltiple de las técnicas de discretización para cada clasificador y **tests post-hoc**.
- Directrices de [Dem06, GH09].

	FRIEDMAN	IMAN-DAVENPORT	HOLM
NB	Rechaza H_0 (0,029)	No necesario	Control = DIF_10
AODE	Rechaza H_0 (0,041)	No necesario	Control = DIF_5 • DIA_10
HAODE	Acepta H_0 (0,052)	Rechaza H_0 (0,049)	Control = DIF_10 • Fel

- La **hipótesis nula** (H_0) establece que **no existe diferencia entre algoritmos**.
- En la 3ª columna, el método tomado como control por el test de Holm y marcado con • los **métodos con los que se halló diferencia** estadística.
- En todos los casos $\alpha = 0,05$.
- (Entre paréntesis el p-valor obtenido en cada caso).



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

Estudio de los diferentes clasificadores:

- **Test de Friedman:** diferencia estadística en **todos los casos**.
- **Tests de Holm y Nemenyi:**
 - **Holm:** HAODE es significativamente mejor que NB y AODE en todos los casos.
 - **Nemenyi:** sólo discrepa en que HAODE no es mejor que AODE cuando se discretiza con Fel.
- En todos los casos el **ranking** realizado por el **test de Friedman** **coloca a los clasificadores en el mismo orden**, independientemente del método de discretización.



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias



- Componentes del error:

- Sesgo**: error sistemático en el aprendizaje del algoritmo.
- Varianza**: variación aleatoria en los datos de entrenamiento y del comportamiento aleatorio en el aprendizaje del algoritmo (**sensibilidad de un algoritmo**).
- El **error irreducible**: error de un algoritmo óptimo (el **nivel de ruido en los datos**).

Motivación

Clasificadores basados en Redes Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE (HAODE)

Métodos de discretización

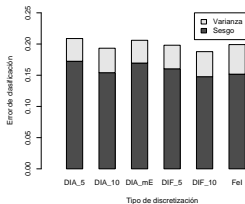
Metodología experimental y resultados

Estudio de la tasa de acierto

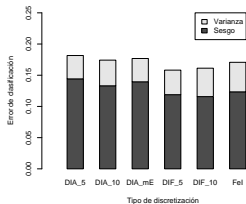
Estudio del sesgo y la varianza

Conclusiones y trabajo futuro

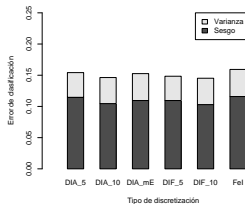
Referencias



(a) NB



(b) AODE



(c) HAODE

Outline

1 Motivación

2 Clasificadores basados en Redes Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE (HAODE)

3 Métodos de discretización

4 Metodología experimental y resultados

Estudio de la tasa de acierto

Estudio del sesgo y la varianza

5 Conclusiones y trabajo futuro



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

Conclusiones y trabajo futuro

- Estudio del efecto en términos de **tasa de acierto**, **sesgo** y **varianza** de los tipos de **discretización** más utilizados hasta ahora en los **clasificadores NB, AODE y HAODE**.
- **Objetivos:**
 - 1 Averiguar si el uso de una determinada técnica de discretización es más beneficiosa que otra para un clasificador.
 - Mejores resultados en media con **DIF_10** para NB y **HAODE** y **DIF_5** para AODE.
 - 2 ¿Puede el uso de un método de discretización **alterar el ranking** de clasificadores?
 - **No**, sólo varía la distancia entre ellos.
- **Trabajo futuro:** estudio de **técnicas alternativas de discretización**, ej. discretización proporcional o división en igual frecuencia sin fijar previamente el número de intervalos ([YW09]).



Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias



Gracias

Motivación

Clasificadores
basados en Redes
Bayesianas

Naive Bayes

AODE

Clasificador Hybrid AODE
(HAODE)

Métodos de
discretización

Metodología
experimental y
resultados

Estudio de la tasa de
acierto

Estudio del sesgo y la
varianza

Conclusiones y trabajo
futuro

Referencias

- [Dem06] Janez Demšar.
Statistical Comparisons of Classifiers over Multiple Data Sets.
J. Mach. Learn. Res., 7:1–30, 2006.
- [GH09] Salvador García and Francisco Herrera.
An Extension on “Statistical Comparisons of Classifiers over Multiple Data Sets” for all Pairwise Comparisons.
J. Mach. Learn. Res., 9:2677–2694, 2009.
- [YW09] Ying Yang and Geoffrey I. Webb.
Discretization for Naive-Bayes Learning: Managing Discretization Bias and Variance.
Mach. Learn., 74(1):39–74, 2009.



- [ZW05] F. Zheng and G.I. Webb.
A Comparative Study of Semi-naive Bayes Methods
in Classification Learning.
*In Proc. of the 4th Australasian Data Mining Conf.
(AusDM05), pages 141–156, Sydney, 2005.*
University of Technology.

