

Projeto de disciplina

IN1102 - Aprendizagem de Máquina

Docente:

Dr. Francisco de Assis Tenorio Carvalho

Discentes:

Cynthia Moreira Maia

Lucas Benevides Viana de Amorim



**Centro de
Informática**
UFPE

Problema



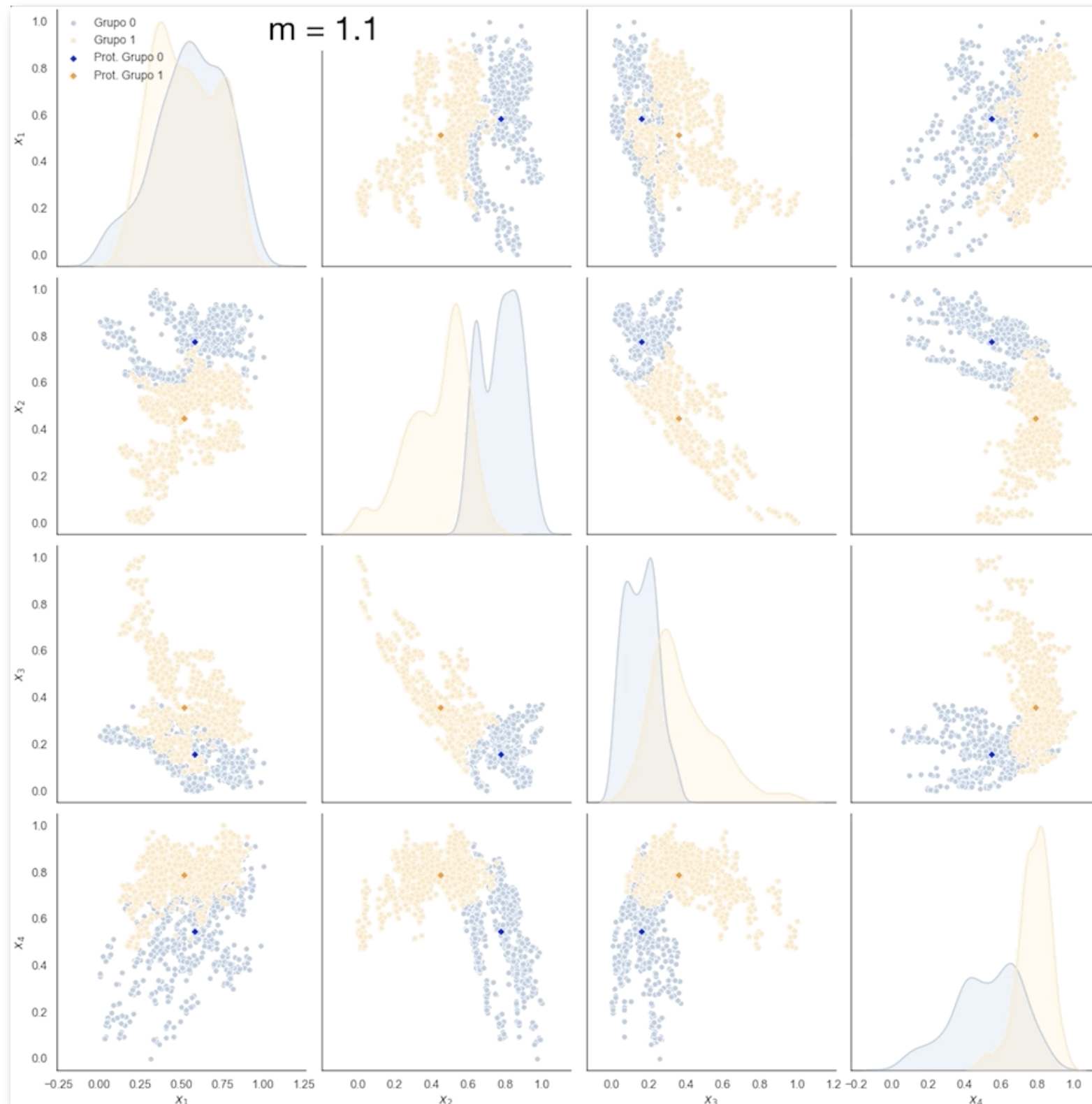
- Dados do *Banknote Authentication Data Set (UCI)*
 - Identificar cédulas entre genuínas e forjadas.
 - Atributos obtidos automaticamente de imagens:
 - Variância (x_1), Assimetria (x_2), Curtose (x_3), Entropia da imagem (x_4).
 - Duas abordagens para o reconhecimento: Supervisionada e não-supervisionada.

Experimento 1

Descrição

- Algoritmo de agrupamento fuzzy VKFCM.
 - *Kernelização* da métrica.
 - Pesos automaticamente definidos para os atributos.
 - Distância adaptativa local.
- Executamos 100 vezes para 3 diferentes valores de m .
- Seleccionamos o melhor para cada m .
- Calculamos métricas para agrupamento *hard* e *fuzzy*.

Resultados e discussão



Protótipos dos dois grupos para os diferentes valores de m .

| m | Grupo | x_1 | x_2 | x_3 | x_4 |
|-----|-------|-----------|-----------|-----------|-----------|
| 1.1 | 0 | 0.5828148 | 0.7767222 | 0.1575695 | 0.5476241 |
| | 1 | 0.5149677 | 0.4500529 | 0.3588664 | 0.7892637 |
| 1.6 | 0 | 0.5905118 | 0.7698198 | 0.1589473 | 0.5563664 |
| | 1 | 0.5068148 | 0.4444541 | 0.3662078 | 0.7894565 |
| 2.0 | 0 | 0.5961270 | 0.7627088 | 0.1632641 | 0.5748662 |
| | 1 | 0.5013098 | 0.4497731 | 0.3610325 | 0.7869952 |

- Variação em $m \rightarrow$ Pouca alteração nos protótipos
- \rightarrow Pouca alteração nas partições *crisp*
- Alterações significativas nos graus de pertinência (a seguir).
- Análise segmentada:
 - Fuzzy
 - Hard

Resultados e discussão

Análise do ponto de vista de agrupamento *fuzzy*

| m | J |
|-----|--------------------|
| 1.1 | 1548.297973916541 |
| 1.6 | 1374.3780335872975 |
| 2.0 | 1127.7134873957498 |

$$J = \sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m \varphi^2(x_k, v_i)$$

| m | MPC | PE |
|-----|---------------|---------------|
| 1.1 | 0.9553 | 0.0379 |
| 1.6 | 0.5737 | 0.3503 |
| 2.0 | 0.3145 | 0.5198 |

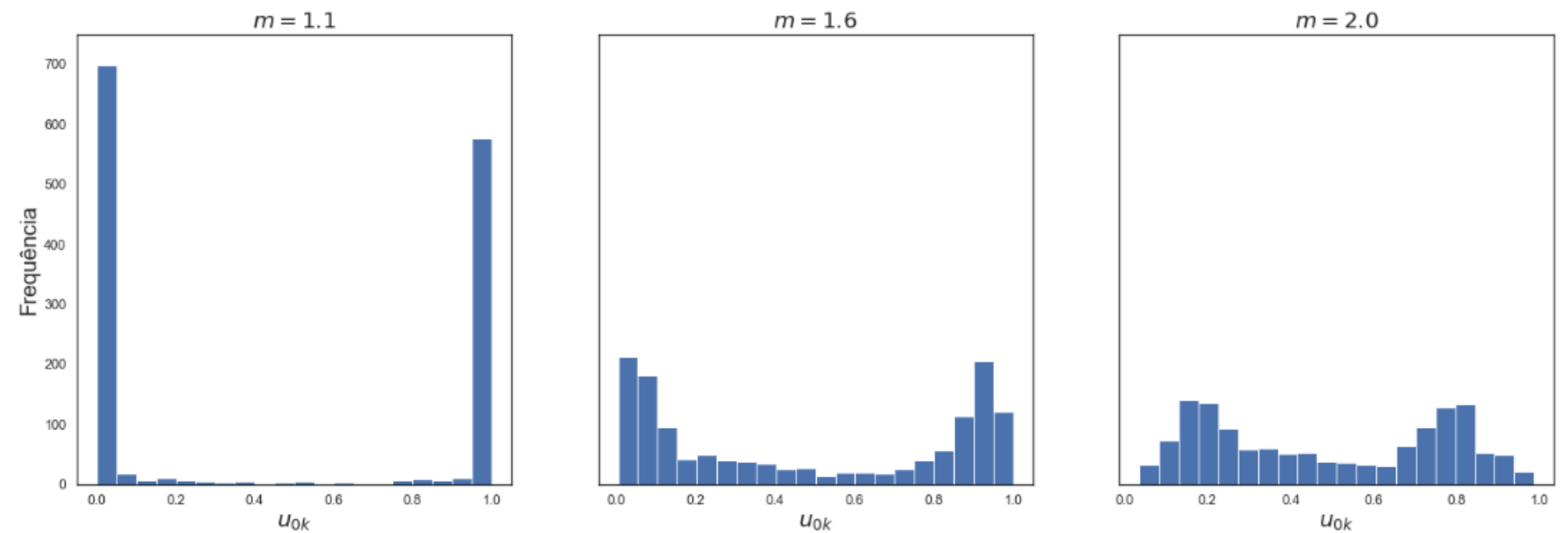


Figura 4: Histogramas para os graus de pertinência dos objetos do grupo 0 para os diferentes valores de m .

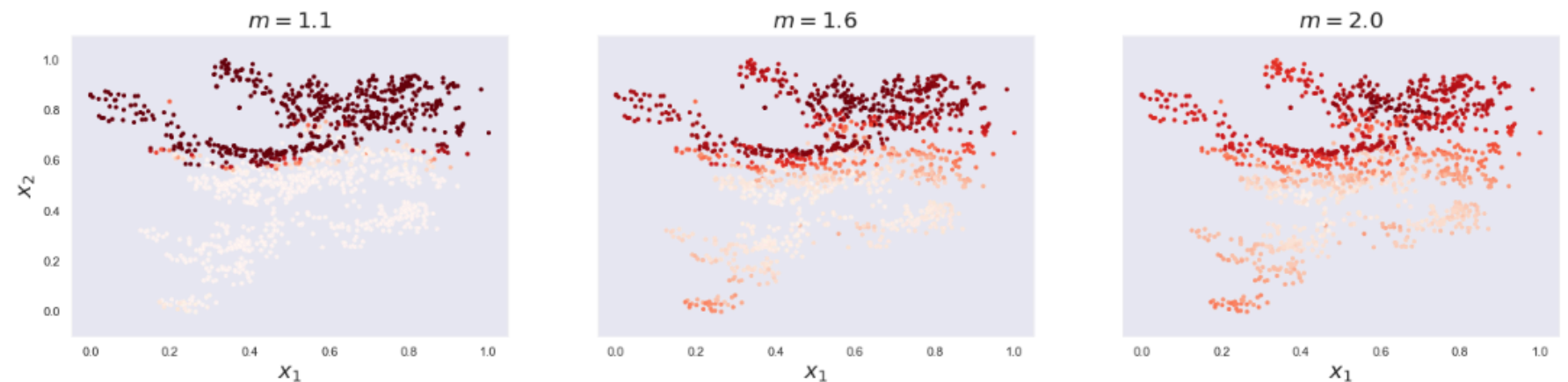


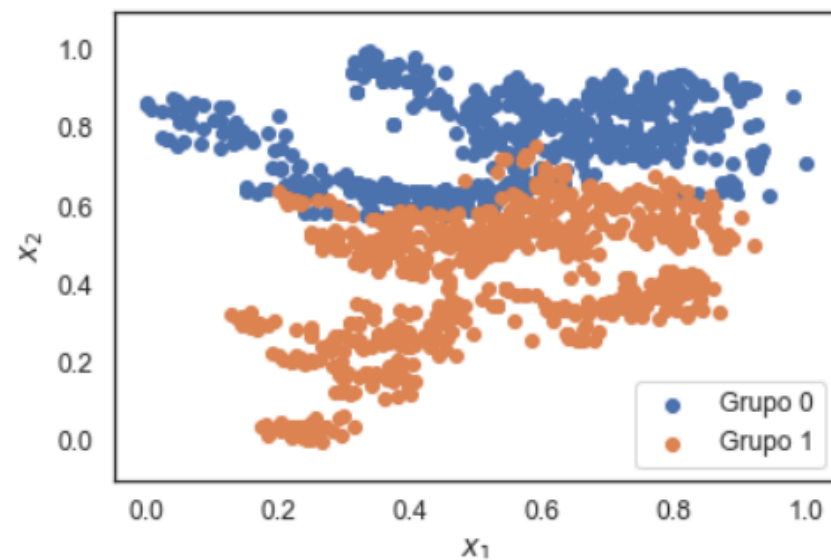
Figura 5: Gráficos de dispersão das instâncias do conjunto de dados, para os diferentes valores de m , com o grau de pertinência ao grupo 0 representado pela escala de cores.

Resultados e discussão

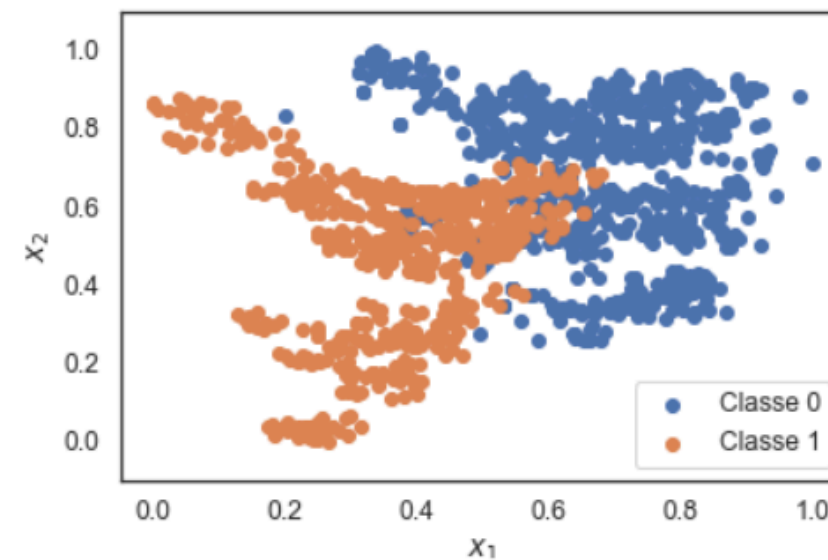
Análise do ponto de vista de agrupamento *hard*

| m | Grupo | Núm. de objetos |
|-----|-------|-----------------|
| 1.1 | 0 | 618 |
| | 1 | 754 |
| 1.6 | 0 | 633 |
| | 1 | 739 |
| 2.0 | 0 | 648 |
| | 1 | 724 |

| m | F-measure | Erro de classificação | Índice de Rand Corrigido |
|-----|---------------|-----------------------|--------------------------|
| 1.1 | 0.5833 | 0.4169 | 0.0268 |
| 1.6 | 0.5816 | 0.4191 | 0.0254 |
| 2.0 | 0.5797 | 0.4213 | 0.0240 |



(a) Grupos



(b) Classes

Figura 9: Gráfico de dispersão das instâncias em que as cores representam em (a): os grupos resultantes (com $m = 1.1$) e em (b): os rótulos das classes originais dos dados.

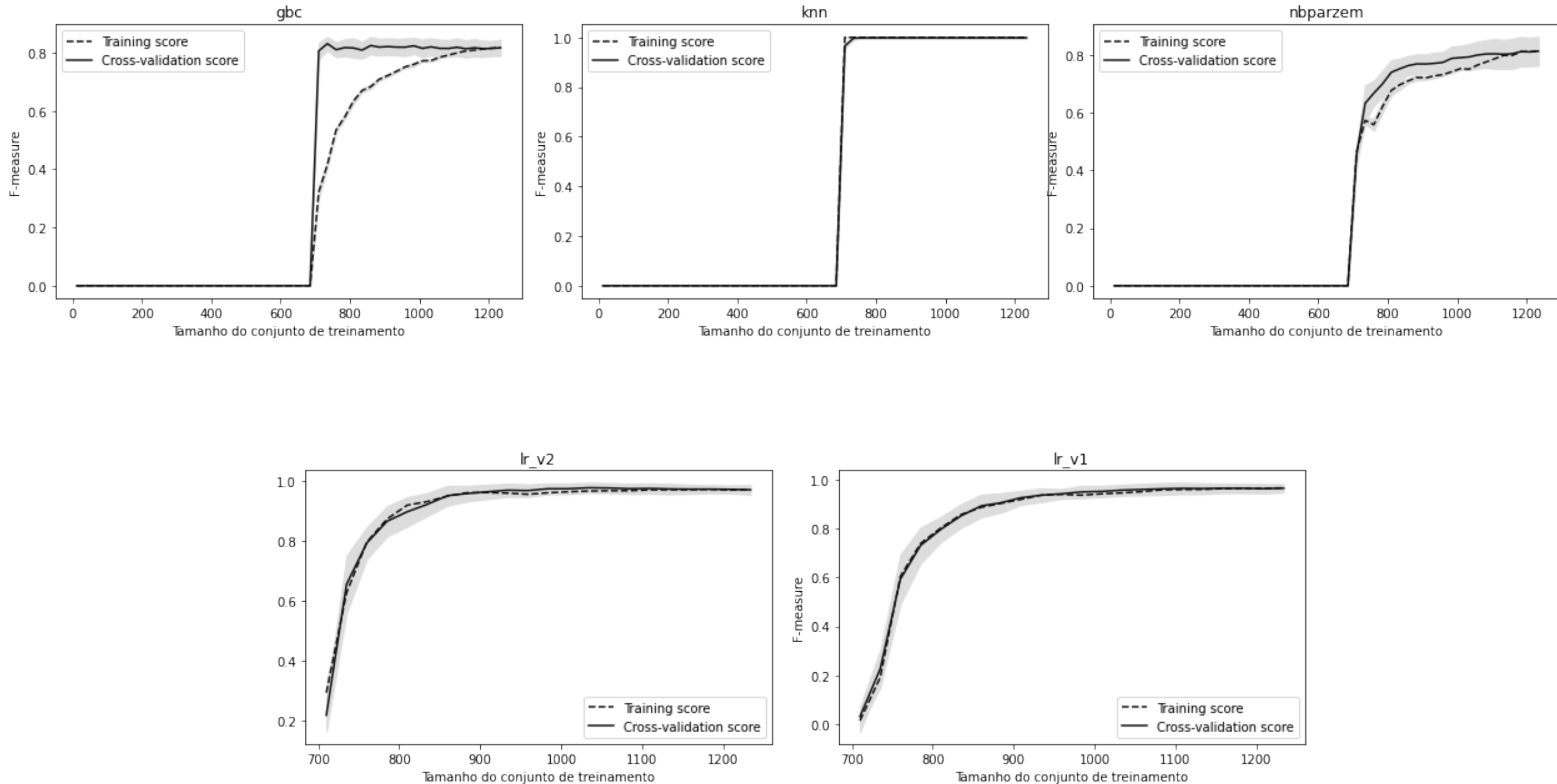
Experimento 2

Descrição

- Comparar performance de classificação (10-fold cv):
 - Bayesiano Gaussiano
 - Bayesiano k-vizinhos
 - Bayesiano janela de Parzen
 - Regressão Logística
 - Regressão Logística - c/ acréscimo de atributos;
 - Ensemble
- Analisar com métricas e teste de Friedman e Nemenyi.

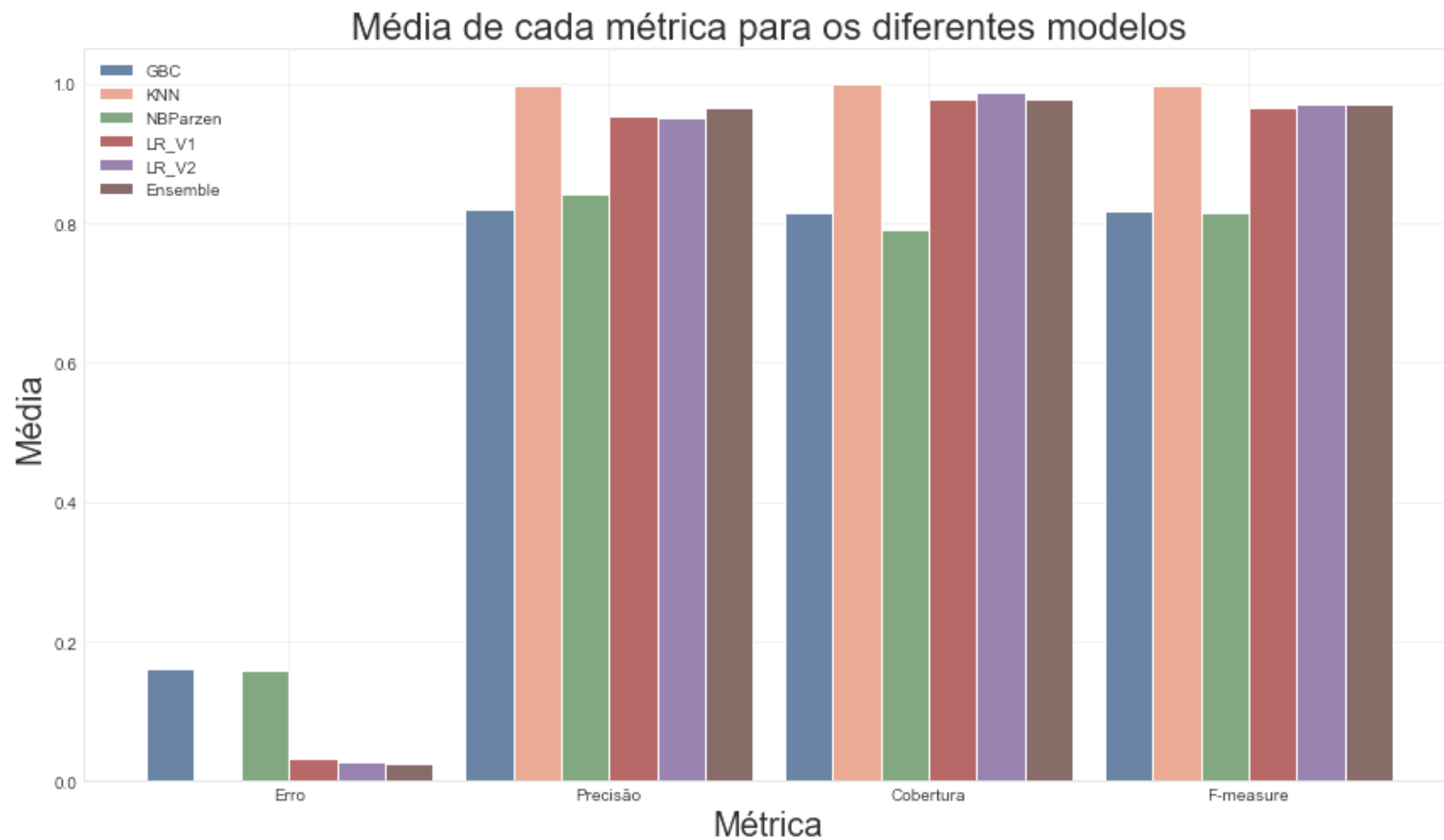
Resultados e discussão

Curvas de aprendizagem



Resultados e discussão

Métricas de performance de classificação



Resultados e discussão

Teste de Hipótese de Friedman

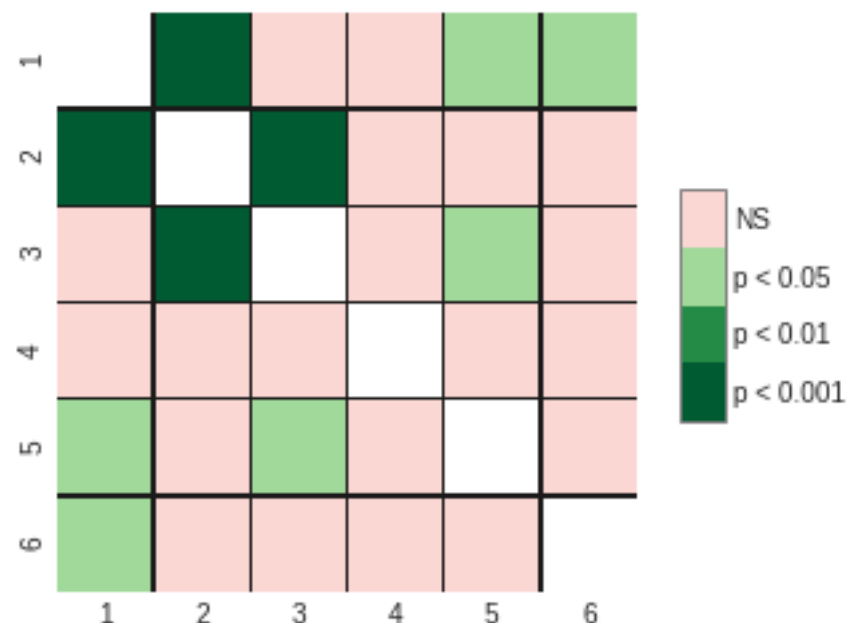
H_0 - Não há diferença significativa no desempenho dos classificadores.

H_1 - Há diferença significativa no desempenho dos classificadores.

Resultado: Estatística = 46.108, p-valor = 0.000

Rejeita H_0

Figura: Teste de Nemenyi para métrica F-measure



1. Bayesiano Gaussiano (GBC)
2. Bayesiano baseado em k -vizinhos (KNN)
3. Bayesiano baseado em janela de Parzen (NBParzen)
4. Regressão logística (LR_v1)
5. Reg. logística c/ atributos adicionais (LR_v2)
6. Ensemble com voto majoritário

Conclusão

- Experimento 1: Realizamos com **sucesso** o experimento de agrupamento fuzzy aplicando uma técnica avançada (VKFCM-LP)
 - Destacamos a **importância do parâmetro m** para a distribuição de frequência de u_{ik} e suas implicações.
 - **Baixa performance** do agrupamento hard como abordagem para classificação. **Potencial uso como atributo adicional** para classificação.
- Experimento 2: Realizamos com **sucesso** o experimento envolvendo 6 modelos, sendo um ensemble.
 - Classificador Bayesiano baseado em k -vizinhos apresentou **melhor desempenho**.
 - Os diferentes classificadores tiveram performance **significativamente diferentes**.
 - Em especial: GBC vs. KNN e GBC vs. LR_v1.