

Aprendizado Descritivo

Aula 09 – Aprendizado Descritivo Supervisionado

Professor Renato Vimieiro

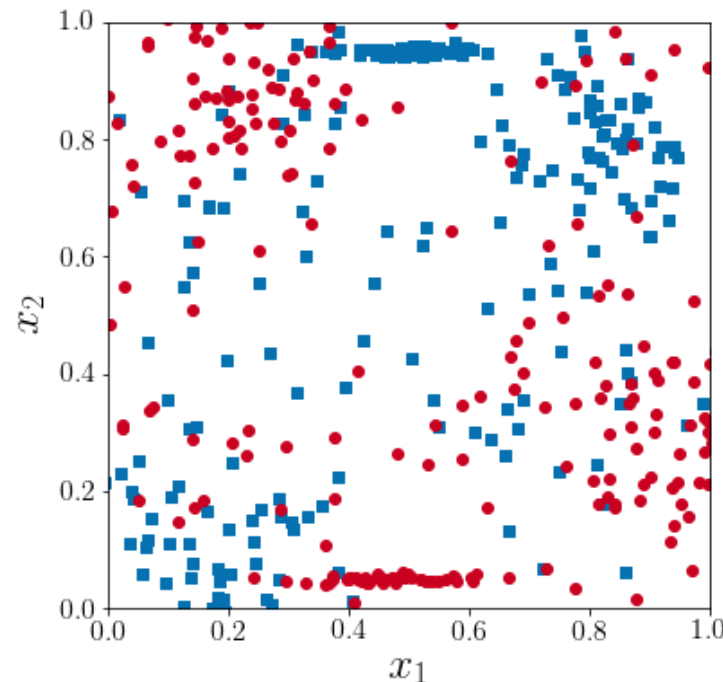
DCC/ICEx/UFMG

Introdução

- Até o momento técnicas de aprendizado descritivo que encontravam padrões em dados não rotulados
- Essas técnicas buscavam regularidades locais nos dados sem a preocupação de aprender o conjunto como um todo
- Por outro lado, técnicas de aprendizado preditivo focam no aprendizado de modelos globais e, para isso, frequentemente fazem uso de dados rotulados
 - Intencionalmente, optam por não aprender pequenas regularidades nos dados
- Nesse sentido, parece existir uma dicotomia entre os dois tipos de aprendizado
- A área de aprendizado descritivo supervisionado se encontra na interseção das duas
 - De uma forma geral, o foco é encontrar regularidades descrevendo dados rotulados

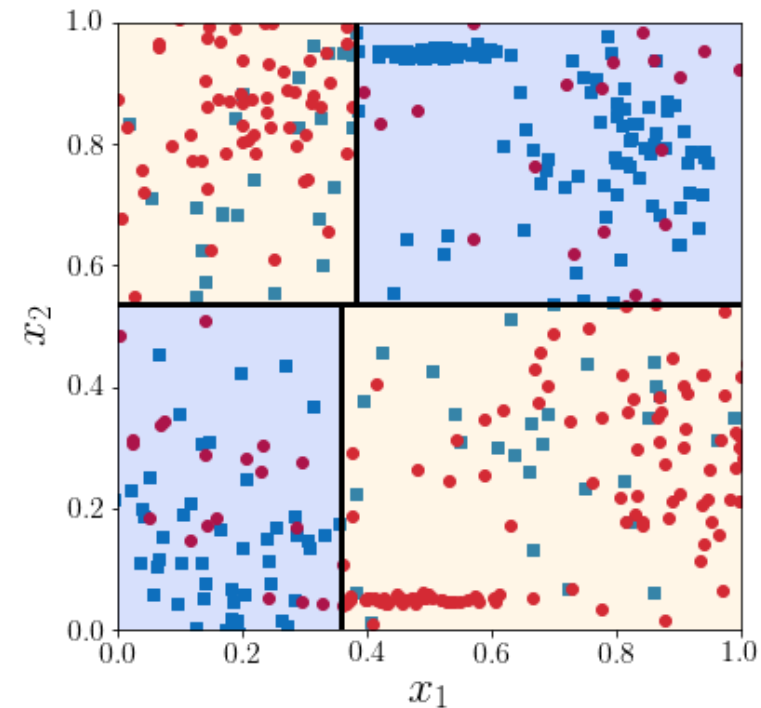
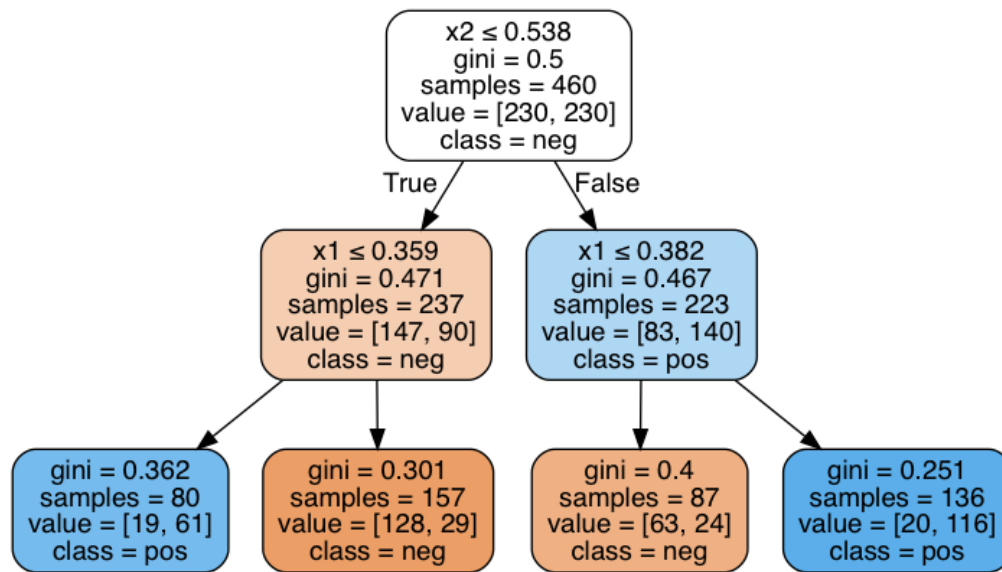
Padrões locais x globais

- Para ilustrar essa diferença entre aprendizado descritivo supervisionado e algoritmos de classificação, consideremos o seguinte conjunto de dados rotulado



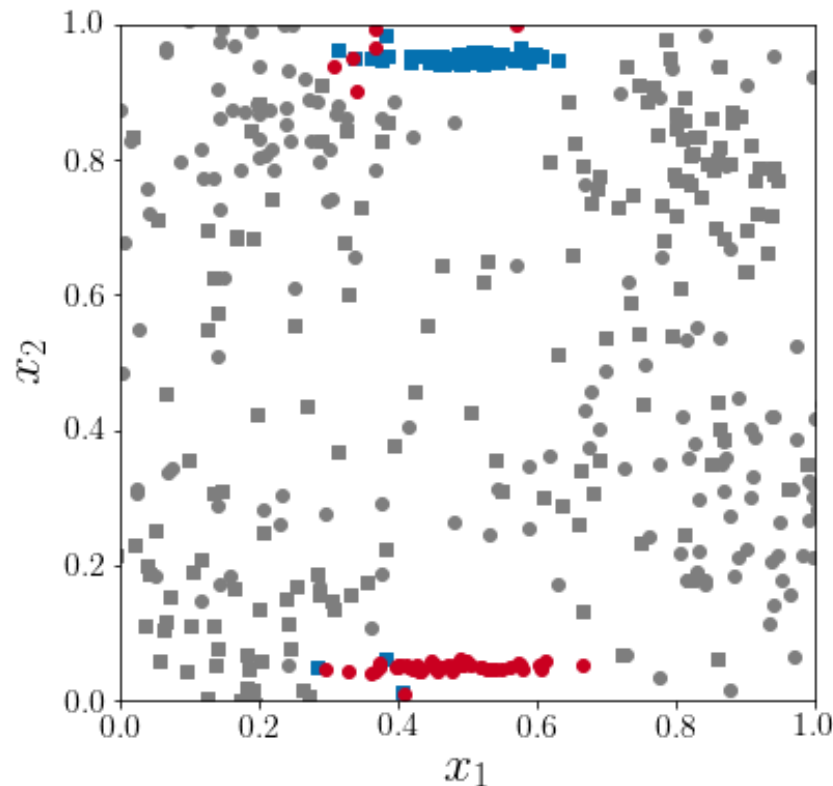
Padrões locais x globais

- Árvores de decisão



Padrões locais x globais

- Padrões descritivos supervisionados



$$\sigma_1 \equiv x_1 \in [0.3, 0.7) \wedge x_2 \geq 0.9$$

$$\sigma_2 \equiv x_1 \in [0.275, 0.7) \wedge x_2 \leq 0.1$$

Aprendizado descritivo supervisionado

- A área de aprendizado descritivo supervisionado evoluiu de forma isolada em três frentes:
 - Padrões contrastantes (contrast set mining)
 - Padrões emergentes (emerging pattern mining)
 - Descoberta de subgrupos (subgroup discovery)
- O que há de comum e que une as três técnicas é o fato delas buscarem padrões locais com distribuições não usuais de alguma característica alvo
 - Normalmente, essa característica alvo é um atributo classe (rótulo das transações)
- Alguns autores referenciam essas técnicas como "descoberta de regras descritivas supervisionadas"
 - Ou Mineração supervisionada de padrões descritivos

Contrast set mining

- Padrões contrastantes foram propostos por Bay e Pazzani (2001) como uma forma de encontrar padrões interessantes em dados rotulados
- Para eles, os objetos continuam sendo caracterizados por atributos categóricos/nominais, porém, agora, divididos em grupos
 - Os grupos são os rótulos dos objetos e são todos disjuntos entre si
- Eles definem que um itemset é um contrast set se existirem pelos menos dois grupos onde a probabilidade de ocorrência diverge

Contrast set mining

- Formalmente:

$$\exists i j \ P(\text{cset} = \text{True} \mid G_i) \neq P(\text{cset} = \text{True} \mid G_j)$$

$$\max_{ij} |\text{support}(\text{cset}, G_i) - \text{support}(\text{cset}, G_j)| \geq \delta$$

- A ideia então é buscar padrões com suportes relativamente diferentes entre os grupos que sejam estatisticamente diferentes em confiança
 - Usam teste chi-quadrado para avaliar se as confianças são significativamente diferentes
- Tarefa adaptada para considerar apenas diferença de suporte

Emerging pattern mining

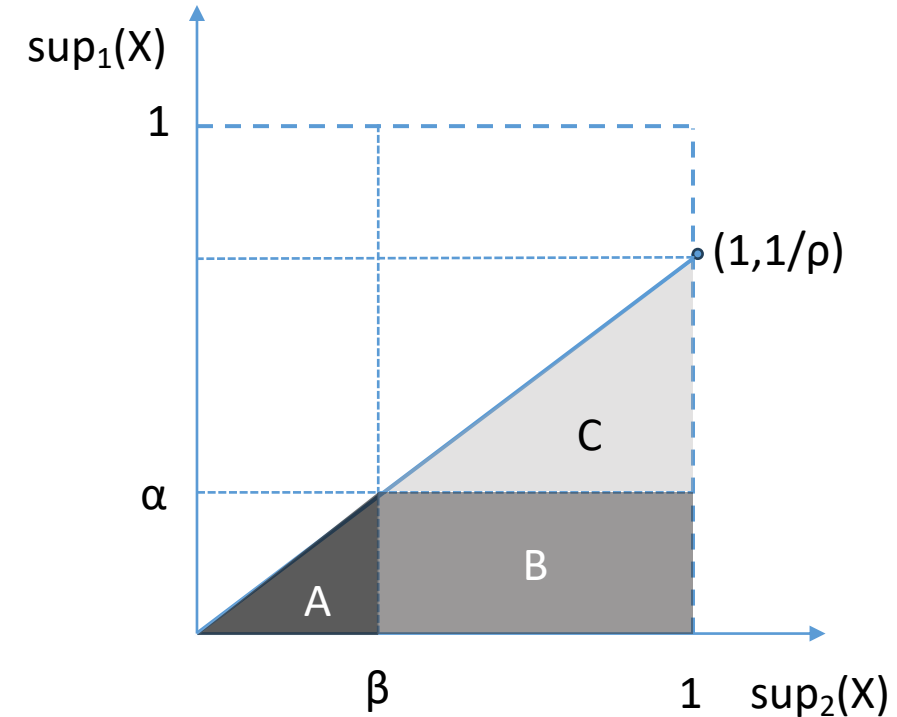
- Emerging pattern mining foi proposta por Dong e Li (1999)
- A ideia dos autores era que seria necessário encontrar padrões que capturassem mudanças entre conjuntos de dados
- Um padrão emergente (emerging pattern) é definido como um itemset cujo suporte aumenta '*significativamente*' de uma base de dados para outra
 - Nesse caso, são considerados explicitamente apenas dois grupos
- A medida de interesse é chamada de *growth rate*, sendo definida como:
 - $$GR(X) = \begin{cases} 0, & \text{se } \sup_1(X) = 0 \text{ e } \sup_2(X) = 0 \\ \infty, & \text{se } \sup_1(X) = 0 \text{ e } \sup_2(X) > 0 \\ \frac{\sup_2(X)}{\sup_1(X)}, & \text{caso contrário} \end{cases}$$

Emerging pattern mining

- Quando se fala em padrões emergentes, assume-se uma direção aonde eles 'emergem'
 - Pelo growth rate, um padrão é infinitamente interessante se ele não ocorre em uma base, mas ocorre pelo menos uma vez na outra
- Assim como na mineração de padrões frequentes, usualmente se estabelece um limiar de crescimento/emergência mínima
 - Um itemset X é um padrão emergente se $GR(X) \geq \rho > 1$
- Podemos elencar diferentes tarefas considerando suporte e growth rate mínimos

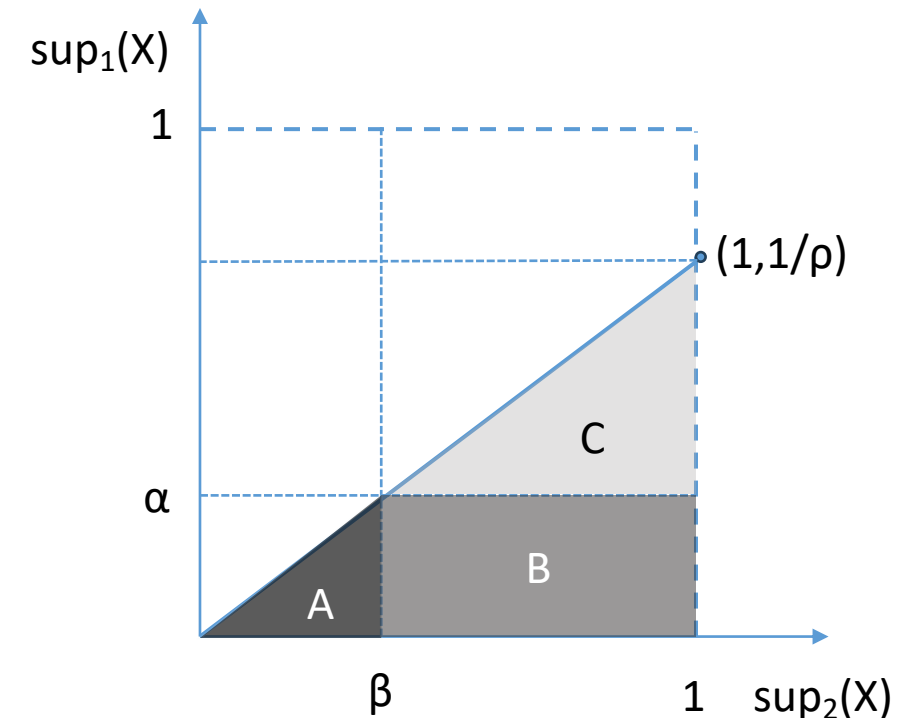
Emerging pattern mining

- A figura ao lado, reproduzida de Dong e Li (1999), define o espaço de busca de padrões emergentes
- Cada itemset é representado por um ponto nesse espaço
 - Se $\text{sup}_1(X) = s_1$ e $\text{sup}_2(X) = s_2$, então X é representado por (s_1, s_2)
- Os padrões emergentes são os itemsets abaixo da reta
- A interseção das retas que passam por α e β é importante
 - $\beta = \alpha \times \rho$
 - O interesse é que esse ponto seja o mais próximo da origem possível



Emerging pattern mining

- Os polígonos A, B e C representam três tipos de tarefas associadas a mineração de padrões emergentes
- O triângulo A são itemsets infrequentes em ambos conjuntos de dados
 - Tarefa: mineração de padrões emergentes raros
 - Difícil execução pelo volume de padrões candidatos
- O triângulo C são itemsets frequentes em ambos conjuntos de dados
 - Tarefa: mineração de padrões emergentes frequentes
 - Possivelmente, trará padrões irrelevantes
 - Pode ser resolvido com pós-processamento nos itemsets frequentes de cada base
- O retângulo B são itemsets infrequentes na base 1 e frequente em 2
 - Tarefa: mineração de padrões emergentes
 - Padrões com maior potencial de relevância prática (i.e. maior aplicabilidade)
 - Pode-se aproveitar algoritmos já desenvolvidos anteriormente



Descoberta de subgrupos

- Entende-se como descoberta de subgrupos a seguinte tarefa (Wrobel 1997):
 - *[...] given a so-called population of individuals (objects, customer, ...) and a property of those individuals we are interested in. The task of subgroup discovery is then to discover the subgroups of the population that are statistically "most interesting", i.e. are as large as possible and have the most unusual statistical (distributional) characteristics with respect to the property of interest*
- Assim, compara-se a distribuição do atributo alvo no subgrupo e em seu complemento (ou população), verificando se elas são significativamente diferentes
 - Se o atributo alvo é a classe/rótulo dos objetos, então compara-se o suporte das classes no grupo e complemento, verificando se são diferentes
 - No caso de rótulos binários, verifica-se o suporte de uma das classes no grupo e complemento

Descoberta de subgrupos

- O objetivo da tarefa então é encontrar padrões que discriminem as transações desses dois subconjuntos
- Os padrões podem ser itens como no caso de mineração de itemsets frequentes
 - Porém é mais comum se definir uma linguagem desses padrões
- A linguagem dos padrões é a conjunção de predicados definidos sobre os atributos da base
 - Numéricos: $\alpha \geq x$; $\alpha = x \dots$ (e.g. idade ≥ 10)
 - Categóricos: $\alpha = x$; $\alpha \notin X$ (e.g. cor=azul; cargo $\in \{\text{programador, analista}\}$)
 - Padrão: $\alpha \geq x \wedge \beta \in X$ (e.g. salario $\geq 3K \wedge$ cargo $\in \{\text{programador, analista}\}$)

Descoberta de subgrupos

- Duas das métricas de qualidade mais usadas para avaliar os padrões são:

- $Q_g(X) = \frac{sup^+(X)}{sup^-(X)+g}$

- $WRAcc(X) = p(X) * (p(+|X) - p(+))$

- $sup^+(X)$: é o suporte positivo (no subconjunto D^+)
 - $sup^-(X)$: é o suporte negativo (no subconjunto D^-)
 - $p(.)$: é a probabilidade (proporção) da variável

Descoberta de subgrupos

- Note que, ao contrário do suporte, essas métricas não são anti-monotônicas.
- Assim, não podemos usar diretamente os algoritmos de mineração de itemsets frequentes para buscar esses padrões
- Existem alguns tipos de algoritmos para a tarefa:
 - Exaustivos: adaptações dos algoritmos clássicos (Apriori, FP-Growth) como Apriori-SD e SD-Map
 - Heurísticas gulosas: algoritmo beam search como o algoritmo SD
 - Heurísticas evolucionárias: algoritmos genéticos, programação genética, como MESDIF, NMEEF, SSDP
 - Algoritmos de sampling: MCTS4DM, MiSoSouP

Equivalência entre as tarefas

- Teorema: Contrast set mining é compatível ('equivalente') a descoberta de subgrupos
- Prova:

$$\begin{aligned} WRAcc(X) &= P(X) \cdot [P(G_1|X) - P(G_1)] \\ &= P(G_1 \cap X) - P(G_1)P(X) \\ &= P(G_1 \cap X) - P(G_1)[P(G_2 \cap X) + P(G_1 \cap X)] \\ &= P(G_1 \cap X) - P(G_1 \cap X)P(G_1) - P(G_1)P(G_2 \cap X) \\ &= (1 - P(G_1))P(G_1 \cap X) - P(G_2 \cap X)P(G_1) \\ &= P(G_2)P(G_1 \cap X) - P(G_2 \cap X)P(G_1) \\ &= P(G_2)P(G_1)P(X|G_1) - P(G_2)P(G_1)P(X|G_1) \\ &= P(G_1)P(G_2)[P(X|G_1) - P(X|G_2)] \end{aligned}$$

Como o tamanho das bases é constante, maximizar o WRAcc é o mesmo que maximizar a diferença de suporte

Equivalência entre as tarefas

- Teorema: Emerging pattern mining é equivalente a descoberta de subgrupos
- Prova: A definição de growth rate é equivalente à do Q_g com $g = 0$ ■
- Embora não tenha sido demonstrada a equivalência direta entre CSM e EPM, o fato de ambas serem compatíveis com SD nos permite deduzir tal equivalência.
- A tarefa de SD ganhou mais notoriedade na literatura que as demais. Grande parte dos estudos usa o termo Subgroup Discovery ao invés de qualquer outro
 - Alguns estudos usam a expressão supervised descriptive pattern mining, mas são mais raros
- Por essa razão, focaremos os estudos na tarefa de SD, lembrando que se trocarmos a função de qualidade, retornamos a qualquer uma das outras duas

Leitura

- Novak, P. K., Lavrač, N., & Webb, G. I. (2009). Supervised descriptive rule discovery: A unifying survey of contrast set, emerging pattern and subgroup mining. *Journal of Machine Learning Research*, 10(2).
- Ventura e Luna cap. 1
- Dong e Bailey cap. 1

Aprendizado Descritivo

Aula 08 – Aprendizado Descritivo Supervisionado

Professor Renato Vimieiro

DCC/ICEx/UFMG