

Problema da falta de rótulos

Transcrição

Queremos saber o comportamento dos clientes em relação ao uso de cartão de crédito, isto é:

- Valor gasto
- Número de transações
- Limite total
- Frequência de compras

Trata-se de uma pergunta abrangente, diferente de classificar o risco dos clientes em três categorias, isto é, os labels.

Neste caso, não temos labels, então não podemos usar o método de classificar os clientes em categorias distintas de comportamento.

Voltaremos ao exemplo de Pedro Lauro. Suponhamos que o risco de Pedro Laura atrasar o pagamento da fatura é baixo, e queremos descobrir seu comportamento.

É neste ponto que utilizamos a clusterização, que faz parte do aprendizado não supervisionado. Neste ponto poderíamos executar um algoritmo de clusterização e criar grupos. Fariamos de maneira que cada cluster possuía clientes com padrões de comportamento similares.

Suponhamos que depois da execução do algoritmo tenhamos chegado a duas categorias, cada uma abarca diferentes clientes. Essas categorias são "comportamento 1" e "comportamento 2".

Mas como podemos saber as características desses dois comportamentos? E como, ainda, podemos saber se as informações fazem sentido real e validar a técnica que utilizamos?

É justamente a interpretação de clusters que iremos analisar ao longo deste curso.