

Introdução

A clusterização é um método de aprendizado não supervisionado que organiza dados em grupos por similaridade. No entanto, em análises de concorrência, abordagens rígidas falham ao forçar os produtos de uma determinada marca a pertencer a um único agrupamento, ignorando que produtos complexos frequentemente ocupam zonas de transições e competem em múltiplas frentes simultaneamente.

Essa simplificação oculta concorrentes híbridos e áreas de atrito vitais para posicionamento estratégico. Ao não calcular o grau de pertencimento, perde-se a visão de quais produtos estão exercendo pressão em diferentes nichos, resultando em um mapeamento de mercado incompleto e impreciso para tomada de decisão.

Para superar essa limitação, este trabalho adota o agrupamento probabilístico (*Soft Clustering*). A metodologia identifica nuances e sobreposições, permitindo que cada produto seja avaliado por seu grau de afinidade com múltiplos perfis simultaneamente. O resultado é um mapa de vizinhança que revela qual a dinâmica de disputa e as intersecções entre *players* do setor.

Metodologia

Para a segmentação dos produtos e análise de suas variáveis, foram adotadas duas abordagens metodológicas.

K-Means Clássico (Clusterização Rígida):

Agrupa os dados em k clusters exclusivos. O algoritmo minimiza a variância dentro dos clusters (inércia), calculando a distância euclidiana entre cada ponto x e o centróide μ_i do cluster S .

A função objetivo é dada por:

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$$

função objetivo

nº de clusters

nº de casos

caso 1

centróide do cluster j

função de distância

O processo ocorre iterativamente através dos seguintes passos:

- Definição de k .
- O algoritmo inicia selecionando aleatoriamente k pontos para serem os centróides iniciais.
- Cada ponto (produto) é atribuído ao centróide mais próximo, medido pela distância euclidiana.
- O centróide é recalculado como a média de todos os pontos recém-atribuídos a esse cluster.
- O processo de atribuição dos pontos e atualização dos centróides é iterativo até que não ocorram mudanças significativas nas posições dos centróides (estabilização).

Soft K-Means (ou Fuzzy C-Means):

Diferente da abordagem clássica acima, a clusterização suave permite que os produtos possuam um grau de pertencimento a múltiplos clusters simultaneamente. O algoritmo minimiza a função objetivo considerando um parâmetro de suavidade m (onde $m > 1$):

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2$$

grau de pertencimento

A otimização segue a mesma lógica iterativa do método clássico, porém com adaptações baseadas na ponderação:

- Além do número de clusters, define-se o fator de suavização m . Inicia-se uma matriz de graus de pertinência (u_{ij}) aleatória, onde a soma das pertinências de um ponto é igual a 1.
- Diferente da média simples da etapa 3 acima, a posição dos centróides c_j é calculada pela média ponderada pelo grau de pertinência (u_{ij}) de cada ponto. Na etapa 2, atribui-se aleatoriamente os graus de pertinência iniciais (u_{ij}^m) para cada produto em relação a cada cluster.
- Ao invés de uma atribuição rígida, recalcula-se a matriz u_{ij} baseada na distância atualizada de cada produto aos novos centróides, repetindo até a estabilização.

Resultados

A análise de 33 produtos sob 25 variáveis de composição permitiu testar a eficácia da clusterização em um cenário real e multivariado.

- A clusterização rígida, ao forçar a exclusividade de grupos, pode alocar inadequadamente produtos situados em regiões vizinhas. Sem o cálculo do grau de pertencimento, perde-se a precisão sobre a afinidade real de cada produto. Na figura ao lado, os produtos externos às elipses ilustram esse problema de fronteira, onde a transição entre grupos não é capturada pelo modelo convencional.

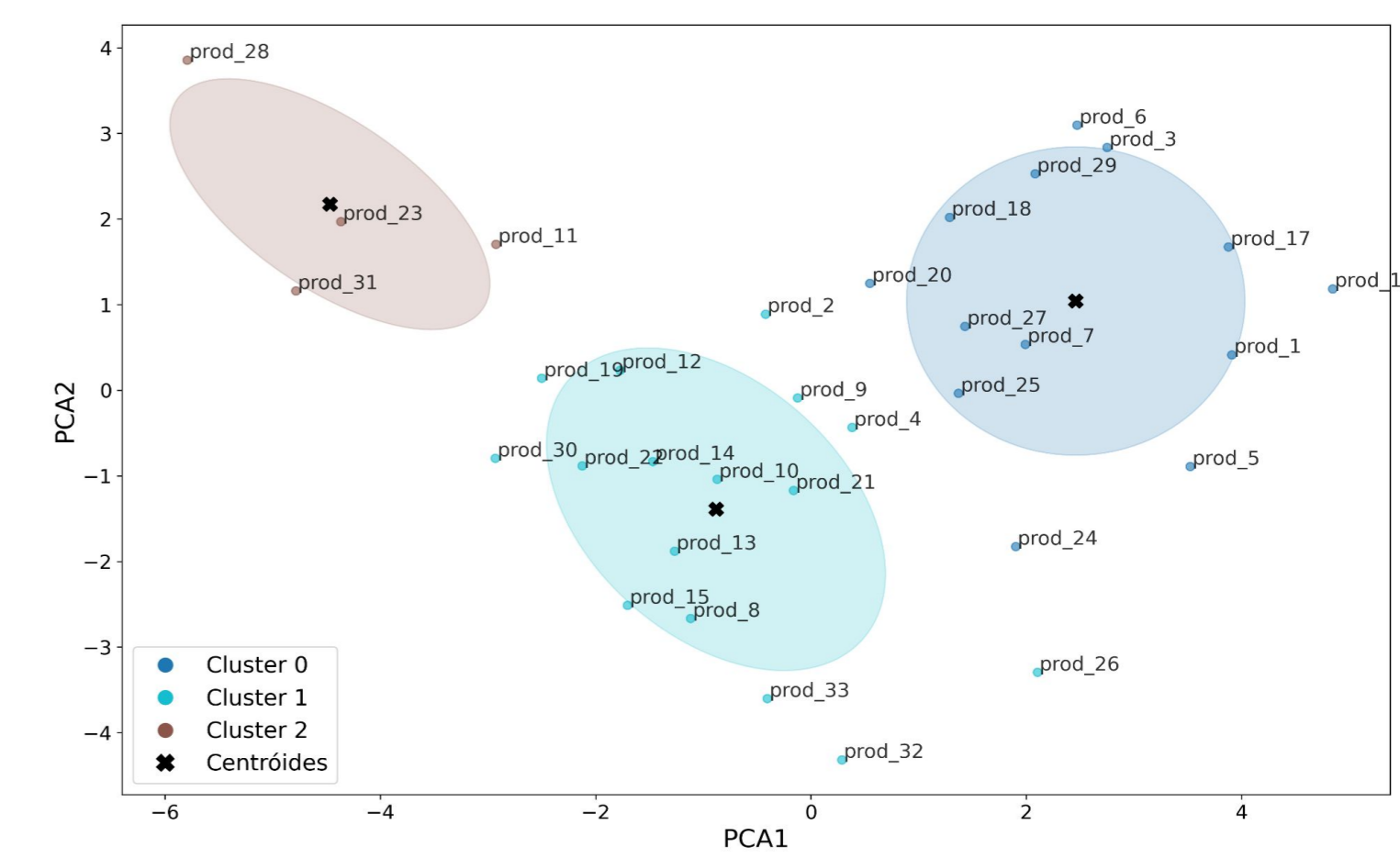


Figura 1: Disposição dos produtos em clusters.

Como a clusterização rígida falha ao ignorar produtos com características mistas, aplicou-se o *K-Means Soft*. Diferente do método tradicional, que força uma escolha única, esta técnica utiliza a lógica difusa para calcular o grau de pertencimento de cada produto em relação aos agrupamentos.

- Conforme demonstrado nas zonas de intersecção da figura ao lado, ao calcular o grau de pertencimento de cada produto para múltiplos grupos simultaneamente, o algoritmo identifica com precisão os perfis híbridos situados nas regiões de fronteira. Essa transição suave entre os clusters elimina as alocações inadequadas do modelo convencional, transformando a ambiguidade dos dados em uma estrutura de inteligência muito mais granular e fiel à realidade dos produtos analisadas.

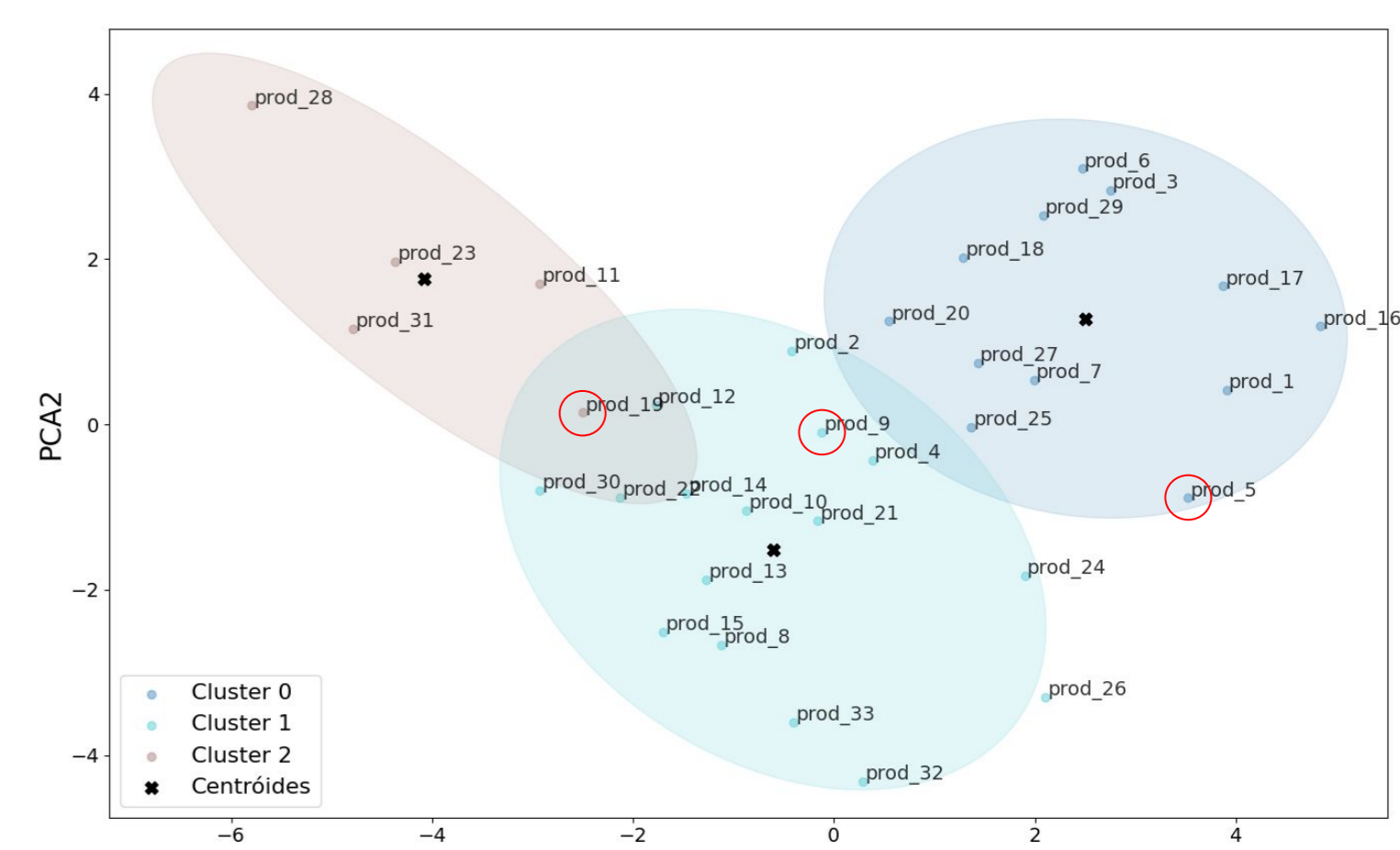


Figura 2: Disposição dos produtos em clusters.

Produto	P(C0)	P(C1)	P(C2)	Cluster
5	70,41%	25,41%	4,17%	0
9	28,14%	58,88%	12,97%	1
19	8,81%	41,11%	49,97%	2

Figura 3: Tabela com graus de pertencimento.

Conclusões

O *Soft K-means* se mostrou ser uma técnica mais completa para a análise de concorrência, pois permite visualizar o mercado não como blocos isolados, mas como ecossistemas de afinidade. Ao identificar o grau de pertencimento de cada produto a diferentes clusters, a metodologia evidencia a proximidade entre competidores que, em modelos de segmentação rígidos, poderiam ser classificados de forma equivocada como distantes.

Dessa forma, essa abordagem possibilita identificar produtos que compartilham o mesmo "núcleo" de características, favorecendo a definição de estratégias defensivas de *market share* mais precisas e eficazes. Além disso, permite localizar concorrentes que transitam entre diferentes nichos (como *premium* ou custo-benefício), revelando potenciais ameaças ou oportunidades de expansão que modelos tradicionais tendem a não capturar.

Bibliografia

- BISHOP, Christopher M. Pattern recognition and machine learning. New York: Springer, 2006.
- SUGANYA, R.; SHANTHI, R. Fuzzy C-Means algorithm: a review. International Journal of Scientific and Research Publications, v. 2, n. 11, 2012.