

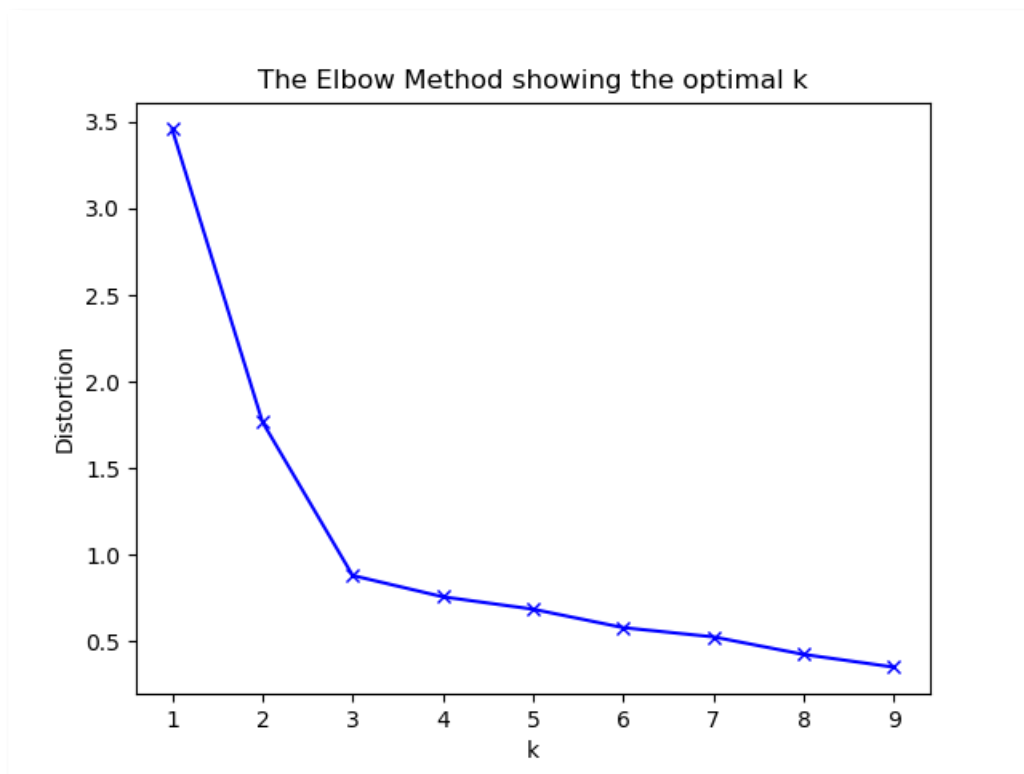
Descrição: esse código escrito em python usa algoritmos de classificação não supervisionada para agrupar produtos recebendo os valores de uma dataset, fornecido, no formato csv.

Dentro do código podemos observar a utilização do algoritmos de agrupamento: Kmeans.

Kmeans : também conhecido como algoritmo de Lloyd, é uma técnica de agrupamento que maximiza dados semelhantes usando a distância do cluster. Uma característica do kmeans, diferente do Meanshift, é que você precisa definir inicialmente o número de cluster antes de iniciar a classificação. Para isso eu usei a técnica do cotovelo (Elbow method).

Obs.: se você não definir o número de cluster, a biblioteca kmeans no scikit learning usado, assume um valor default de 8 clusters.

Elbow method : na prática você utiliza um plano cartesiano (x,y), o número k de cluster e a proporção da dispersão (variância) entre o modelo e o centro do cluster. Para um k inicial, a dispersão decai rapidamente a medida em que os valores de k aumentam, o que fica evidenciado pela curva.



No ponto em que a curva se acentua, $k = 3$ na figura, formando a imagem de um cotovelo, temos um valor de k em que a partir daí valores maiores de k não significam uma melhora significativa nos resultados do agrupamento.

Conclusão: Usando essas técnicas foi possível agrupar os produtos com características semelhantes, e dispor aqueles com características peculiares.

Obs: O código foi criado usando, dentro do anaconda : ipython(5.3.0), scikit-learning (0.18.1), numpy (1.12.1), pandas(0.20.1) e matplotlib(2.0.2). Dentro do código podemos observar a utilização de 2 algoritmos de classificação não supervisionada: Kmeans e MeanShift, ficando a critério escolher.