Centroides são pontos representativos que definem o centro geométrico de um conjunto de pontos em um espaço multidimensional. Em contextos como análise de clusters (clustering), os centroides são frequentemente usados para representar o centro de um cluster.

Em um espaço bidimensional, um centroide seria um ponto cujas coordenadas médias são calculadas em relação a todos os pontos do conjunto de dados. No entanto, esse conceito pode ser estendido para espaços com mais dimensões.

Em algoritmos de clustering, como o K-Means, por exemplo, o centroide de um cluster é atualizado iterativamente durante o processo de agrupamento para garantir que esteja no centro do conjunto de pontos associados a esse cluster. O processo de otimização em algoritmos de clustering visa minimizar a distância entre os pontos do cluster e seu centroide.

A fórmula para o cálculo do centroide em um espaço n-dimensional é dada pela média das coordenadas em cada dimensão. Suponha que você tenha um conjunto de pontos (�1,�1,�1,…,�1),(�2,�2,�2,…,�2),…,(��,��,��,…,��)(*x*1​,*y*1​,*z*1​,…,*w*1​),(*x*2​,*y*2​,*z*2​,…,*w*2​),…,(*xn*​,*yn*​,*zn*​,…,*wn*​). O centroide (��,��,��,…,��)(*Cx*​,*Cy*​,*Cz*​,…,*Cw*​) seria calculado da seguinte forma:

��=�1+�2+…+���*Cx*​=*nx*1​+*x*2​+…+*xn*​​ ��=�1+�2+…+���*Cy*​=*ny*1​+*y*2​+…+*yn*​​ ��=�1+�2+…+���*Cz*​=*nz*1​+*z*2​+…+*zn*​​ ⋮⋮ ��=�1+�2+…+���*Cw*​=*nw*1​+*w*2​+…+*wn*​​

Os centroides são úteis em várias aplicações, desde a análise de clusters até o processamento de imagens e reconhecimento de padrões. Eles fornecem uma representação compacta e central do conjunto de pontos, facilitando a interpretação e a análise dos dados.