

# Product Quantization

Product quantization (PQ) is a technique used in high-dimensional data compression and approximate nearest neighbor search to reduce storage and computational costs. It works by dividing the high-dimensional vector space into smaller subspaces, quantizing each subspace independently, and representing vectors using a combination of quantized codes. These codes are derived from precomputed centroids, enabling efficient storage and distance calculations. PQ is particularly effective in applications like image retrieval and natural language processing, where large-scale datasets require fast similarity searches. By leveraging the reduced representation, PQ balances accuracy and efficiency in managing large datasets.