

K-means clustering for pairs trading

K-means clustering for pairs trading

- Using K-means clustering for pairs trading involves a systematic approach to group stocks into clusters based on similarities in their historical price movements, volatility, or other relevant financial metrics.
- The idea is to identify pairs within these clusters that move in a similar fashion over time, thus making them good candidates for pairs trading.
- Here's a detailed step-by-step guide on how to apply K-means clustering for pairs trading:

Step 1: Data Collection

- Firstly, collect historical price data of the stocks you are interested in. This data typically includes daily closing prices, and possibly, volumes and volatility measures. Data can be sourced from financial databases like Bloomberg, Yahoo Finance, or Google Finance.

Step 2: Feature Selection

- Decide on the features that will be used for clustering. Common features include:
- **Returns:** Calculated as the percentage change in price from one period to the next.
- **Volatility:** Often calculated as the standard deviation of returns over a given period.
- **Trading Volume:** Could be average volume over a period.
- **Other Technical Indicators:** Such as moving averages, RSI, etc.

Step 3: Preprocessing

- **Normalize the data:** Since K-means is sensitive to the scale of the data, ensure that all features are on a similar scale, typically using z-score normalization or min-max scaling.
- **Calculate distances:** K-means clustering involves calculating the Euclidean distance between data points, so all selected features should be numeric.

Step 4: Apply K-means Clustering

- **Choose the number of clusters, K :** The choice of K can be subjective but is crucial. Techniques like the elbow method can be used to determine an appropriate K by plotting the within-cluster sum of squares (WCSS) against the number of clusters and finding the "elbow point" where the rate of decrease sharply shifts.
- **Run the K-means algorithm:** Using a tool or library, such as Python's Scikit-learn, perform the K-means clustering.
- **Analyze clusters:** Review the stocks grouped into each cluster to ensure they exhibit similar financial characteristics.

Step 5: Identify Potential Pairs

- Within each cluster:
- **Select potential pairs:** Look for pairs of stocks that are closely aligned in terms of their price movements and other financial metrics.
- **Assess correlation and cointegration:** For each potential pair, calculate the correlation coefficient and perform cointegration tests to ensure the pairs truly move together and share a long-term equilibrium relationship.

Step 6: Trading Strategy Development

- For each identified pair:
- **Define entry and exit points:** Based on historical data analysis, define when to enter a trade (when the spread between the pair widens beyond a standard deviation) and when to exit (when the spread returns to its mean).
- **Backtest the strategy:** Use historical data to test how the strategy would have performed in the past. This helps in understanding potential returns and risks.

Step 7: Implementation and Monitoring

- **Implement the strategy:** Deploy capital to trade the identified pairs using a broker.
- **Monitor and adjust:** Continuously monitor the performance of the pairs and adjust the parameters as necessary. Reapply K-means periodically to adapt to changing market conditions.

- **Tools and Technologies**

- **Python Libraries:** Use Scikit-learn for K-means, Pandas for data manipulation, NumPy for numerical calculations, and Matplotlib for plotting data.