Machine Learning and Financial Applications

Lecture 7

Statistical arbitrage with machine learning

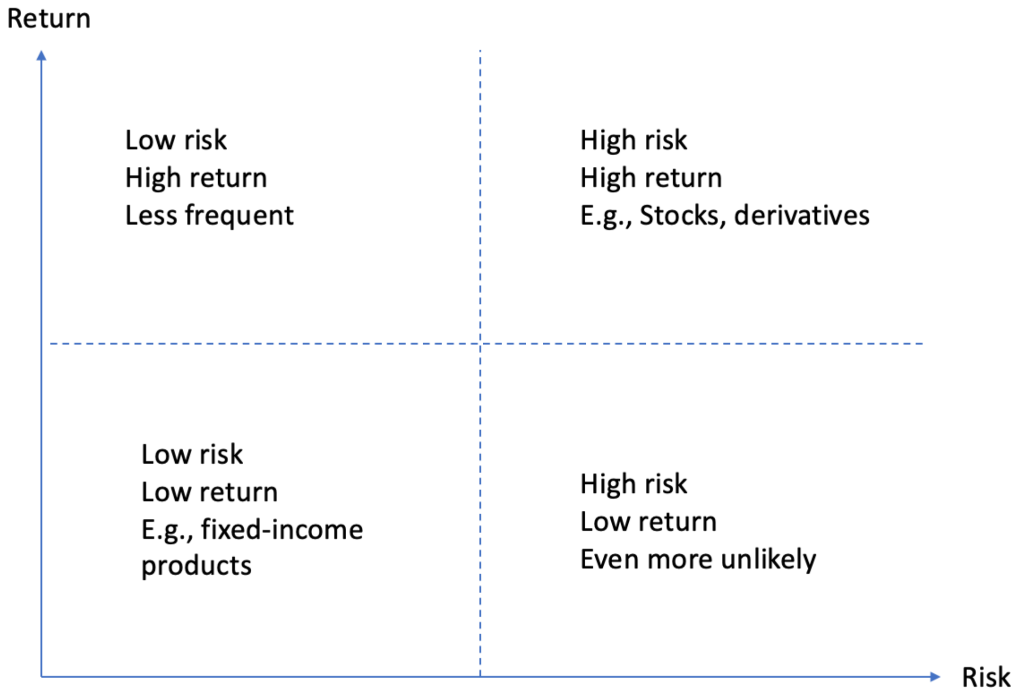
Liu Peng

liupeng@smu.edu.sg

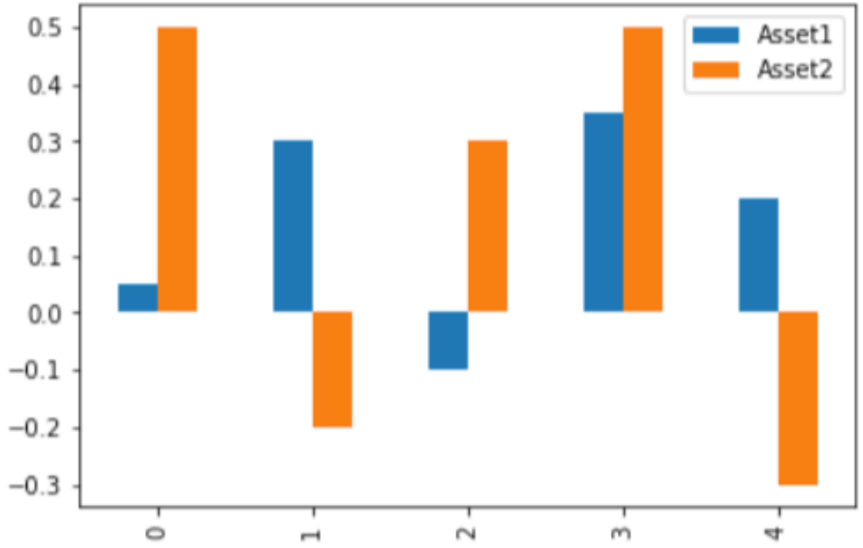
Video tutorial

- Statistical arbitrage and pairs trading https://youtu.be/5Q5nVU0Cdg0
- Cointegration, stationarity, z score in pairs trading https://youtu.be/CbOIYIf0O0s
- Pairs trading with machine learning https://youtu.be/wRAhWpVGCL4
- SVM, random forest, neural network in pairs trading https://youtu.be/PDQ-sg9HoyE
- Implementing the pairs trading strategy https://youtu.be/cDiL3b1HCXA
- Implementing pairs trading strategy using machine learning https://youtu.be/as-fLk2L2xE

Risk and return tradeoff



	Asset1	Asset2
0	0.05	0.5
1	0.30	-0.2
2	-0.10	0.3
3	0.35	0.5
4	0.20	-0.3



Working with Sharpe ratio

- Exercise: build a function or class to assess the Sharpe ratio of trend following strategy on a specific stock and date range for any long and short window sizes
    - Input: stock symbol, date range, window sizes (hint: may not come from the same function)
    - Output: Sharpe ratio

Calculating the risk-adjust return

Return over risk

$$\frac{R_P}{\sigma_P}$$

Sharpe ratio

$$\frac{R_P - R_f}{\sigma_p}$$

Considers the risk-free rate that represents the market benchmark

The numerator is also called the excess return

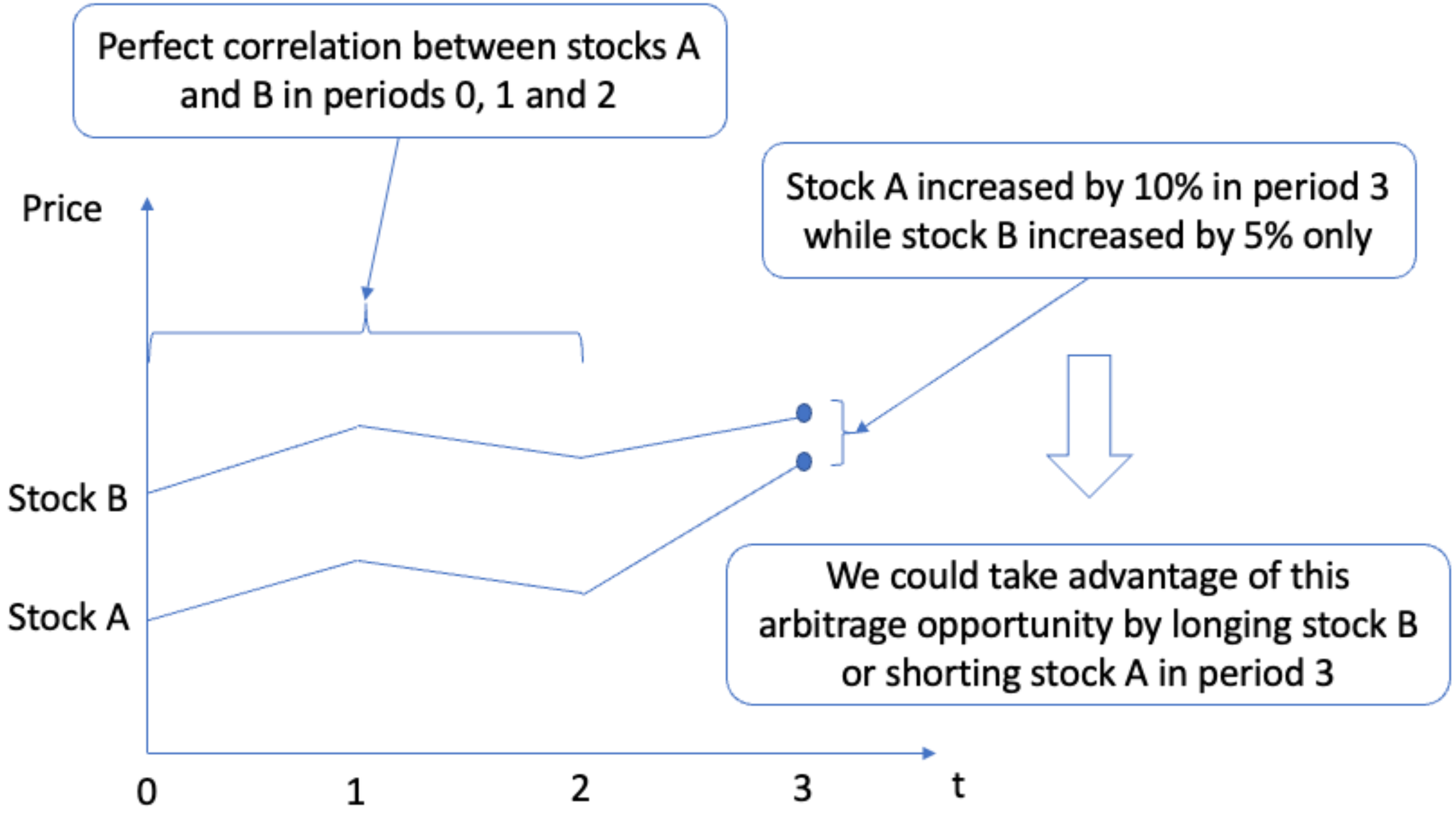
Statistical arbitrage

Use statistical methods to identify statistically significant relationships underlying multiple financial assets and generate trading signals.

A market-neutral strategy that generates profits by taking advantage of temporary market inefficiencies.

Involves two steps: identify pairs of trading instruments based on specific statistical procedures that identify co-movement, and generate trading actions using strategies such as pairs trading.

Statistical arbitrage



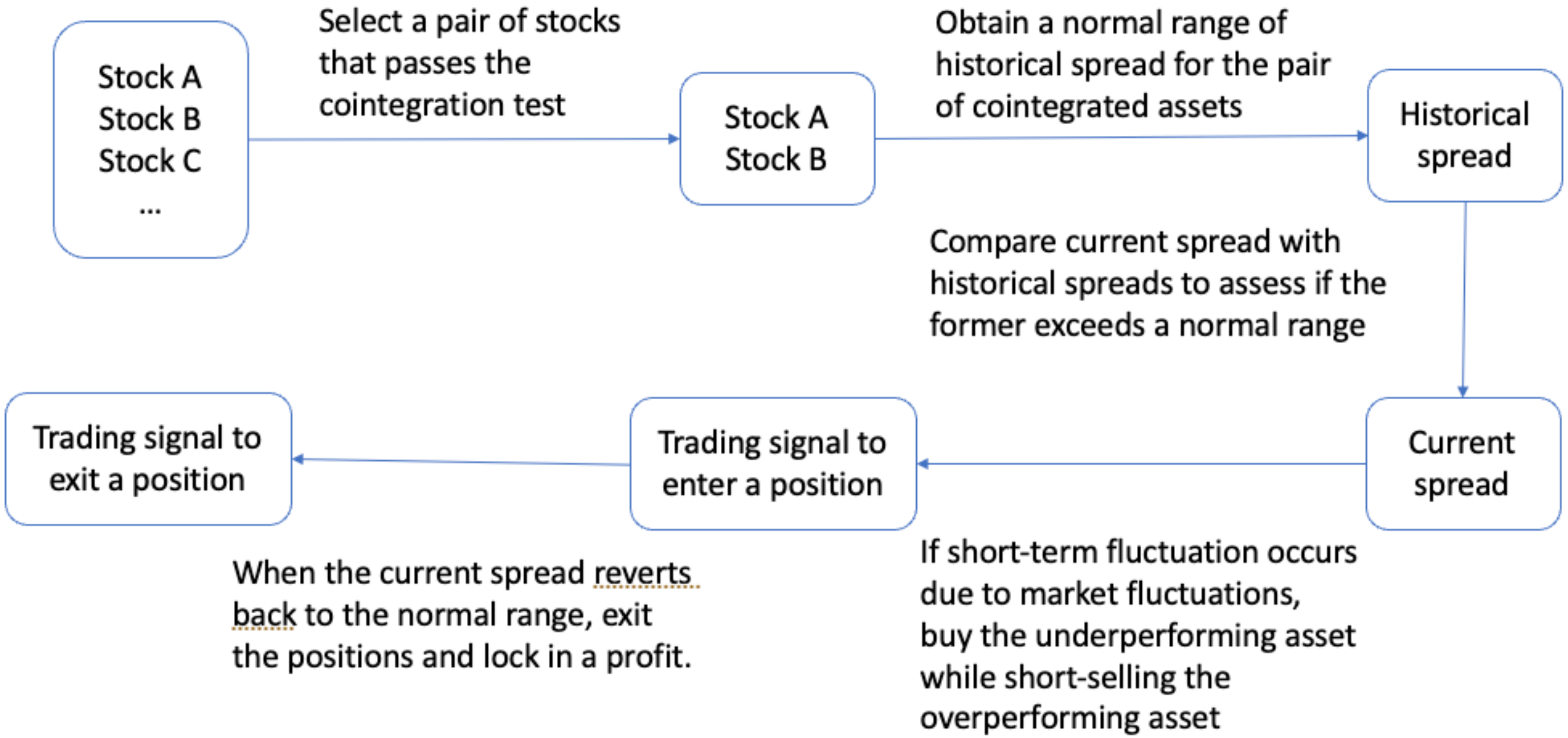
Case study: LTCM

- A hedge fund with billions of AUM and over 50% yearly return
- Well backed by investment experts and academics, including two Nobel prize winners
- Use math to beat the (irrational) market. Model -> Analyze > Forecast.
- Most typical strategy: convergence trade, also called relative value strategy
- Instead of focusing on one asset whose price is often volatile, find two correlated/cointegrated assets and take arbitrage in the case of mispricing due to market fluctuations
- Short-term divergence in asset prices generates trading signals that promise profits due to long-term convergence

Case study: LTCM

- Stay market neural
- Use rigorous statistical tests and stringent thresholds to find highly correlated/cointegrated assets
- Use cheap leverage to amplify profits up to 30 or even 50 times
- Diversify portfolio to control risk by engaging in multiple uncorrelated strategies.
- Even if a single strategy has high volatility, the portfolio will have low volatility due to diversification.
- This assumption crashed when the market crashed

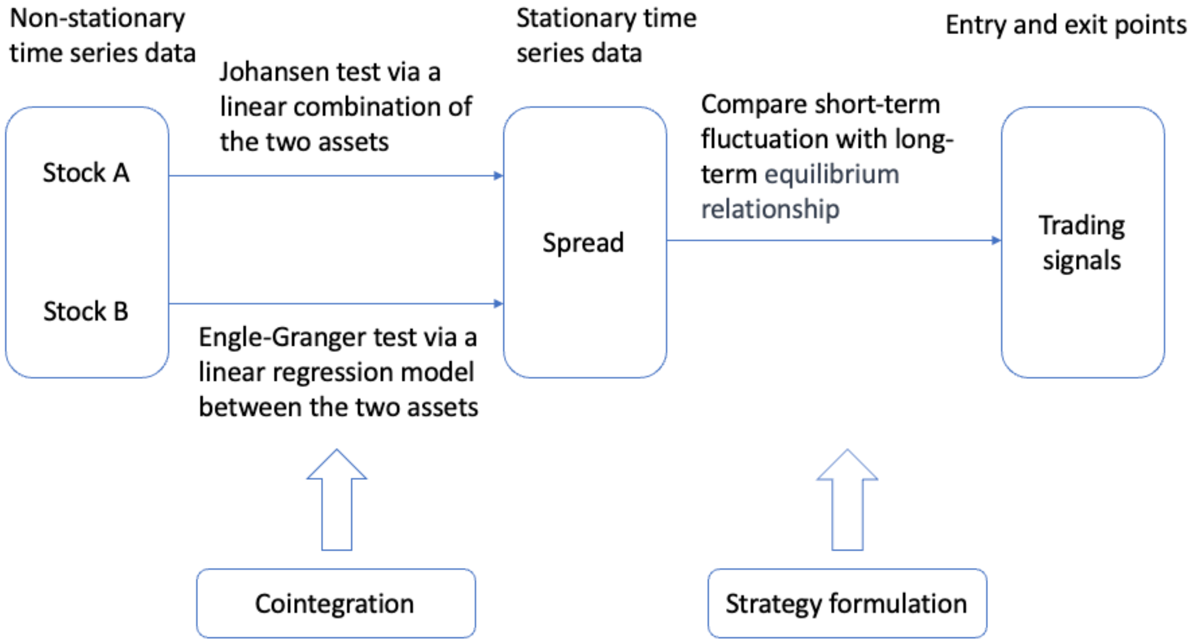
Pairs trading



Cointegration

$$y = \beta_0 + \beta_1 x + \epsilon$$

$$\epsilon = y - (\beta_0 + \beta_1 x)$$



Test for cointegration

- We'll use the Engle-Granger two-step method
  - Estimate the coefficients of the linear regression model between one stock (as the dependent variable) and the other stock (as the independent variable) using ordinary least squares (OLS).
  - Calculate the residuals from the linear regression model.
  - Test the residuals for stationarity using a unit root test, such as the augmented Dickey-Fuller (ADF) test.
  - If the residuals are stationary, the two stocks are cointegrated. If the residuals are non-stationary, the two stocks are not cointegrated.

Stationarity

- A stationary time series is a time series where the statistical properties of the series, including the mean, variance, and covariance at different time points, are constant and do not change over time.

- Normal distribution:

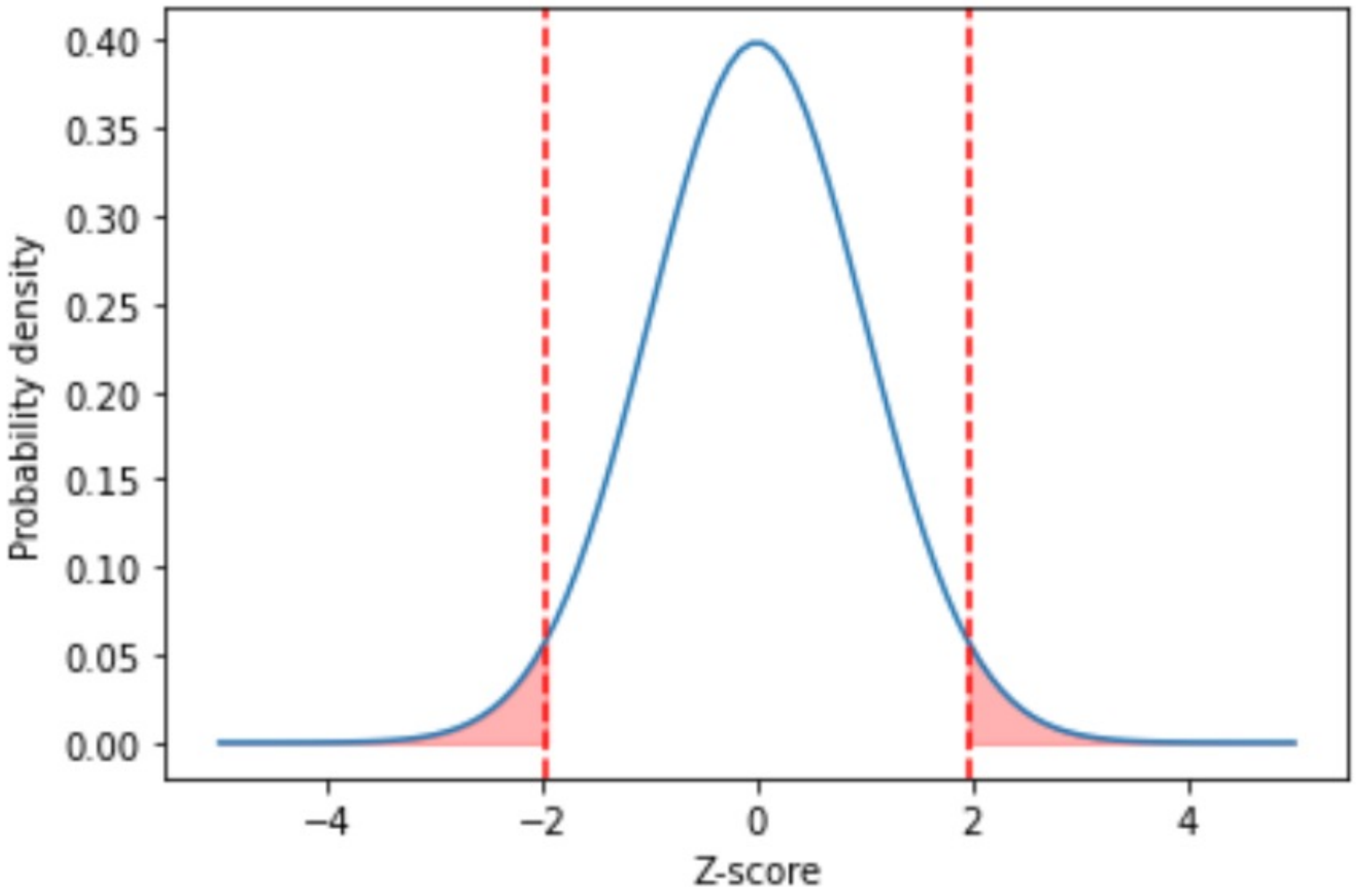
$$y = f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Z score

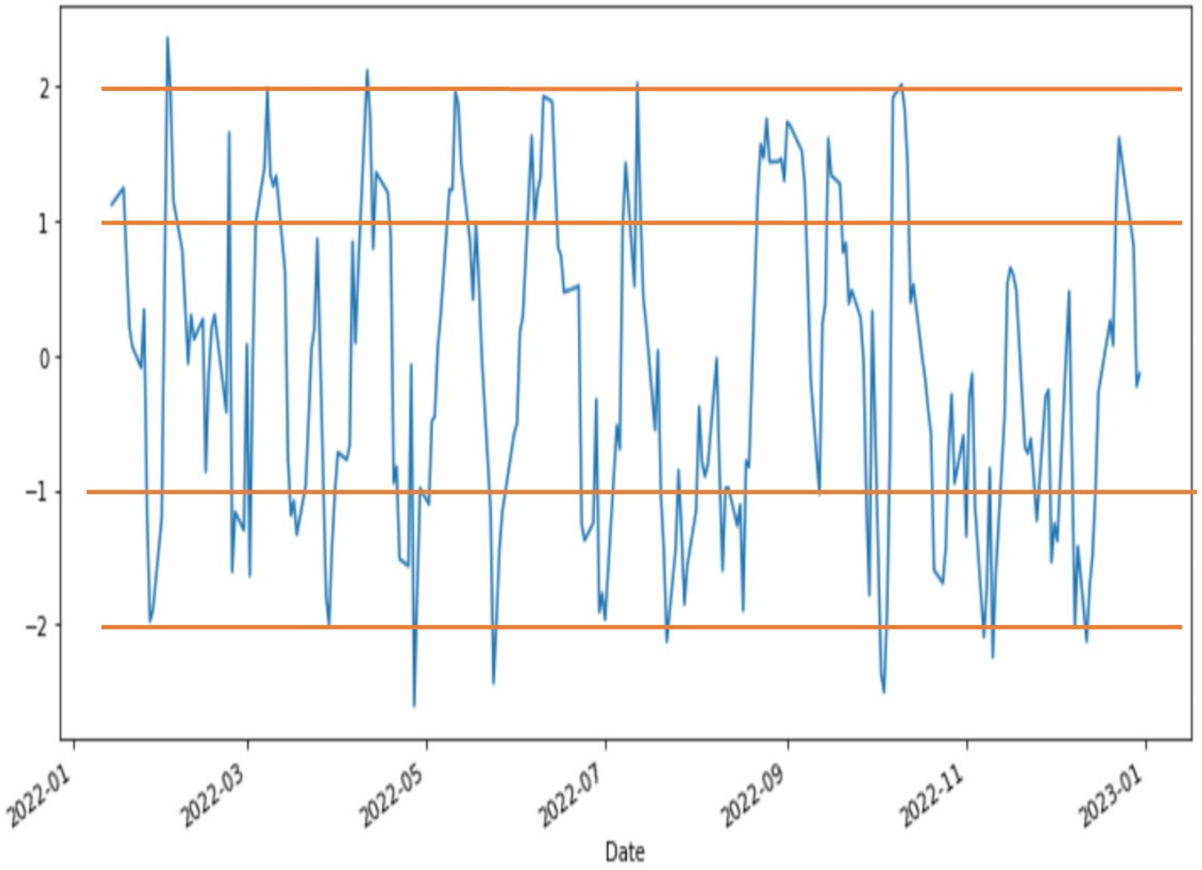
- A measure of how many standard deviations the daily spread is from its mean.

- A standardized score that we can use to compare different distributions

- $z = \frac{x-\mu}{\sigma}$



Implementing pairs trading strategy



If z-score>2, short stock 1 and long stock 2

Maintain current position if 1<z-score<2

If -1<z-score<1, exit positions in both stocks

Maintain current position if -2<z-score<-1

If z-score<-2, long stock 1 and short stock 2

More on pairs trading

Pairs trading is a type of quantitative trading strategy that involves transacting two highly correlated/cointegrated assets at the same time and in the opposite direction.

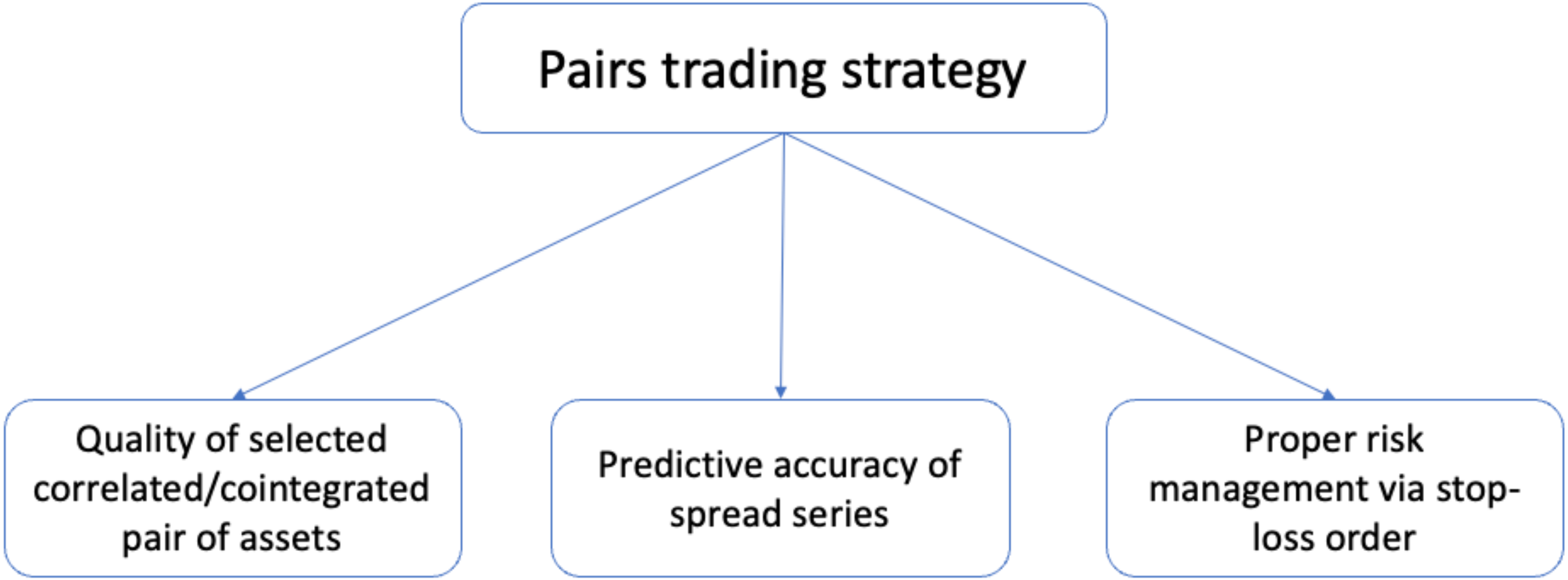
The primary assumption behind pairs trading is that the price spread between two highly correlated or co-integrated assets should exhibit a mean reversion behavior over time.

The two assets identified by the strategy should bear a long-term equilibrium relationship and move in tandem, while any deviation from this pattern is likely to be temporary and will eventually revert back to the mean.

Success of pairs trading

- The success of the strategy depends on two factors: whether the pair of correlated/cointegrated assets with a similar risk profile can be identified, and whether the spread between the two assets can be accurately predicted.
- For example, we used a moving average in the previous chapter to standardize the daily spread into a z-score. Such a moving average acts as the predicted spread, which is then used to compare with the actual spread of the day and derive the unit of deviation in terms of the standard deviation.
- In addition, we also need to have proper risk management in place. When the spread continues to widen and moves in an adverse direction due to unexpected market events, the larger spread can lead to a significant loss.

Success of pairs trading



Machine learning models

- Machine learning models are predictive functions that generate predictions given a specific set of inputs.
- We intend to use a machine learning model in pairs trading to predict the spread between the two assets, which will then be used to identify profitable trading signals.
- Since the spread is a continuous quantity, we will explore regression models in this lecture, including support vector machine (SVM), random forest (RF), and neural network models.
- We will also augment the feature space, i.e., historical spread series, with additional features such as technical indicators.

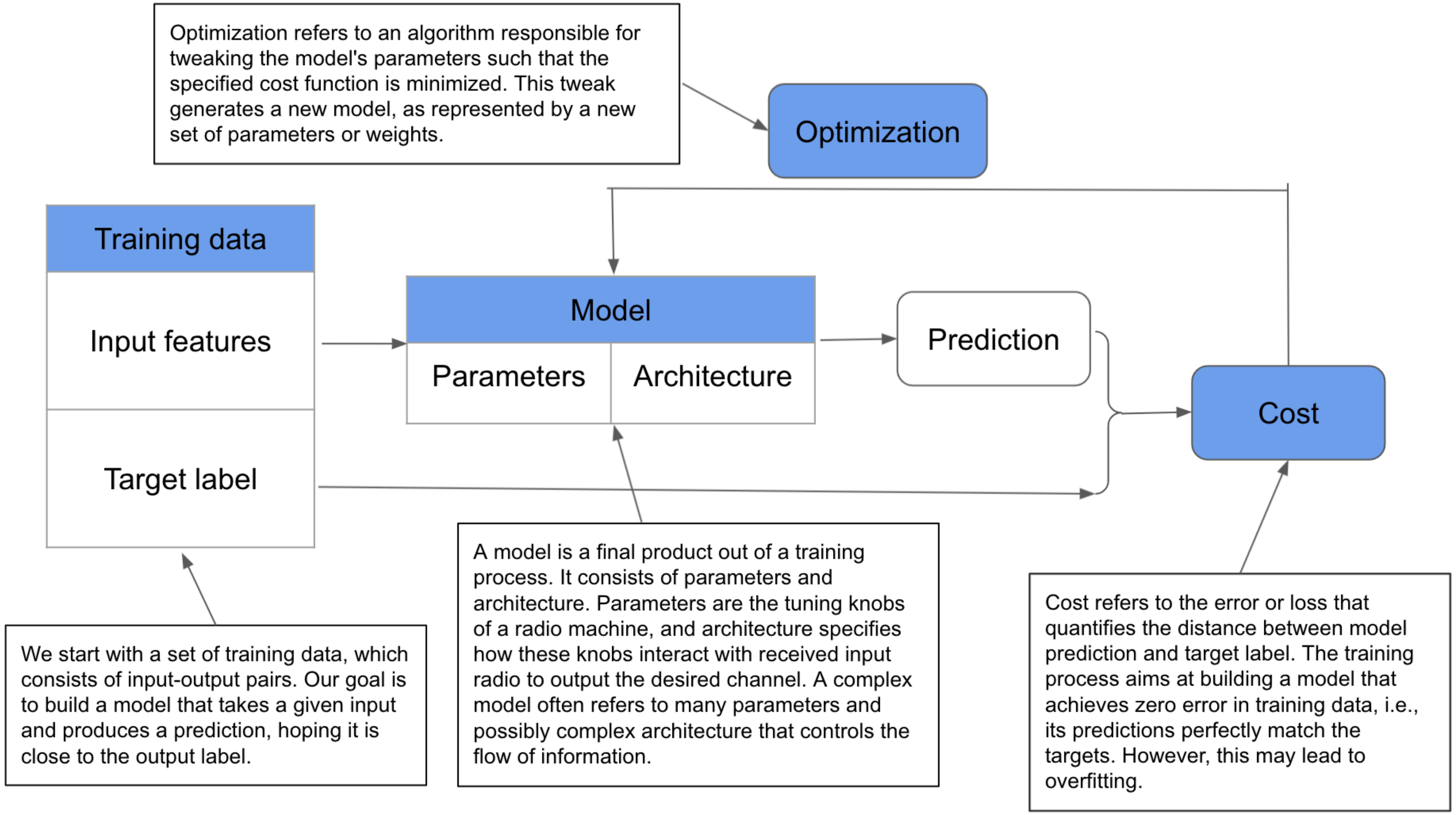
Decomposing a machine learning model

- A typical machine learning model consists of two components: the parameters (or weights) serve as the building blocks of the model, and the model architecture specifies how the input data interact with the parameters to generate the output.
- Model training refers to the process of tuning these parameters such that the model produces a good performance on the test set, and often a relatively good performance on the training set.
- During the training process, the machine learning algorithm adjusts the parameters of the model based on the input data, to improve the accuracy of its predictions on the training data.
- Once the model is trained, it can be used to make predictions for the new data it has not seen before.

Overfitting

- If the model performs too well on the training set but not so well on the test set, then the model is considered as overfitting the training data.
- Since modern models are typically complex in architecture and large in the number of model parameters, overfitting is a common phenomenon in many training situations.
- Proper regularization techniques can be adopted to reduce the chance of overfitting.

Machine learning workflow

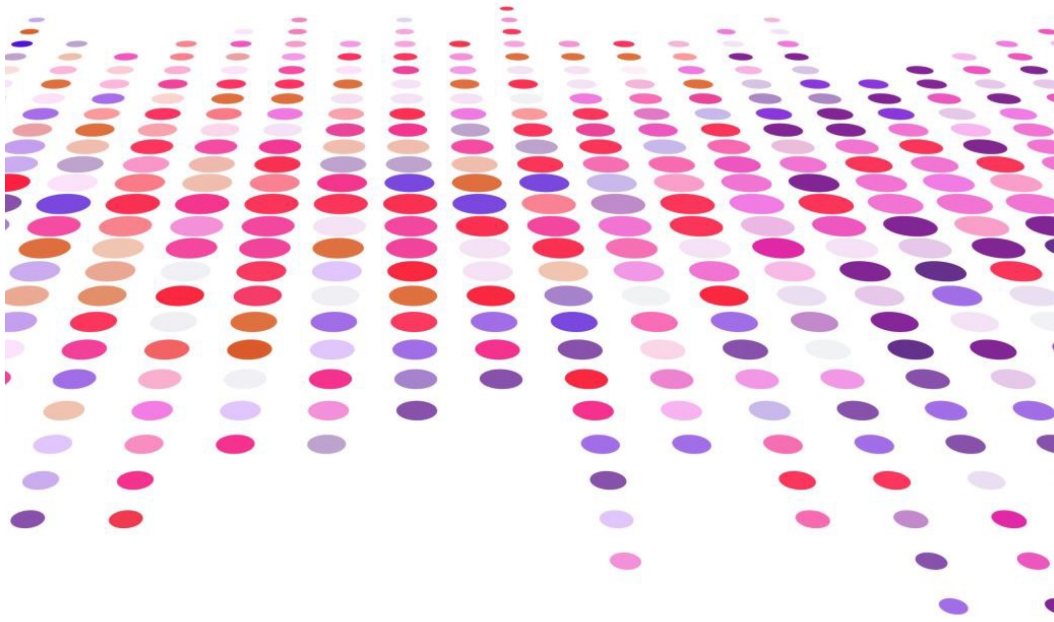


Support vector machine

- Support Vector Machine (SVM) is a popular supervised learning algorithm, especially in the Kaggle community, for both classification and regression.
- In the context of classification, SVM works by mapping the input data from its original feature space into a high-dimensional feature space using a kernel function, and then finding the hyperplane that best separates the different classes of data.
- The hyperplane is chosen in order to maximize the margin between the classes.
- Seeking a boundary based on the principle of maximal margin often leads to a better generalization performance, thus reducing the risk of overfitting.

Support vector regression (SVR)

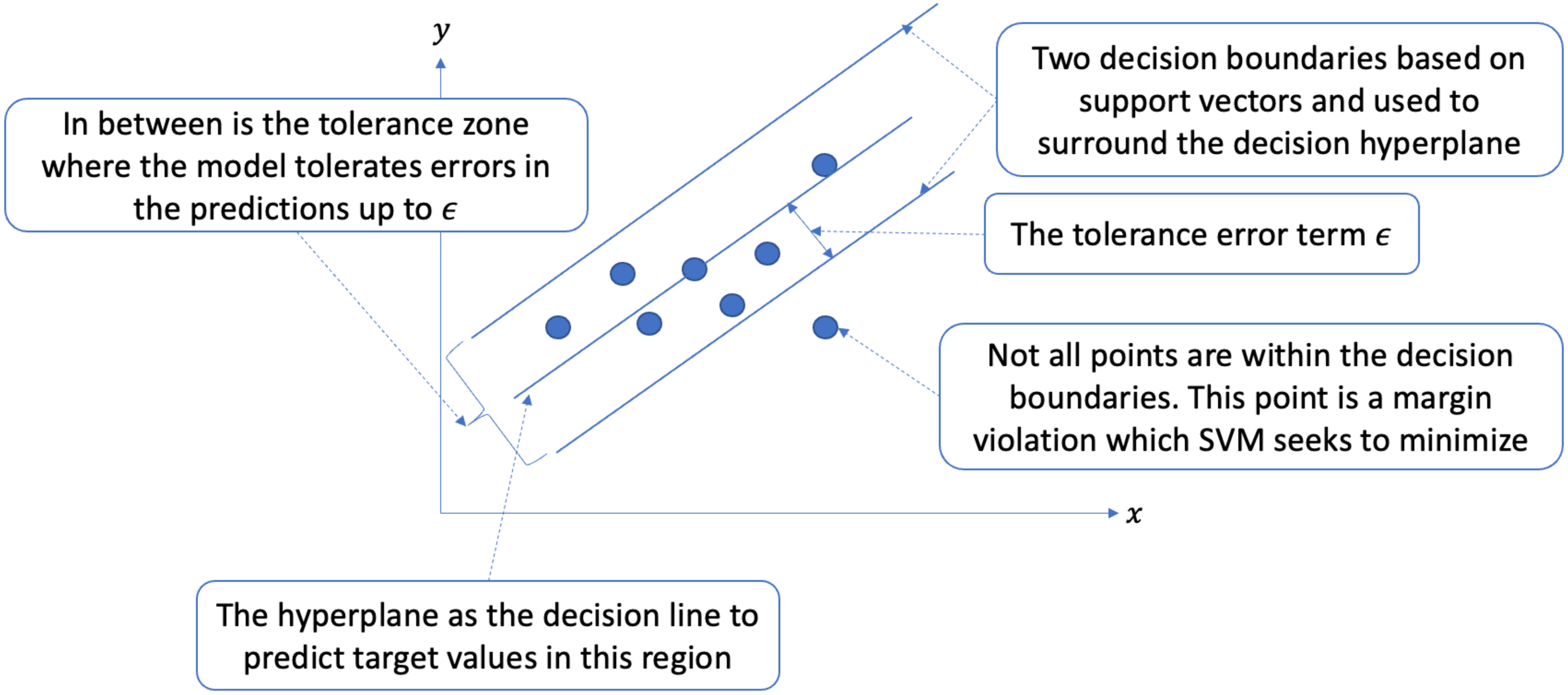
- Since we are interested in predicting the spread as a continuous outcome, making it a regression task, SVM instead finds the hyperplane that best separates the input data while minimizing the margin violations.
- Our goal in the regression task is to fit a hyperplane as closely as possible to the actual data points by minimizing the sum of the squared errors (SSE) as the cost measure between the predicted output and the actual target values.
- Since minimizing SSE toward zero would easily lead to an overfitting model, the SVM model used in regression often assumes an ϵ-insensitive loss function, which allows the model to tolerate some error in its predictions, up to a certain threshold ϵ.



Technical terms in SVM

- A hyperplane is a decision line used to predict the continuous output in the case of regression.
- The data points on either side of the hyperplane within a certain distance (specifically, within ϵ) are called support vectors.
- We can also use these support vectors to draw two decision boundaries around the hyperplane at a distance of ϵ.
- A kernel is a set of mathematical functions that take data as input and transform it into the required form, possibly in a different dimension. These are generally used for finding a hyperplane in the higher dimensional space, which is considered easier to achieve linear separation than finding the same separating hyperplane in the original feature space.

Illustration of SVR



Choice of ϵ

- ϵ controls the tolerance of the margin violation. It determines the trade-off between the model complexity and the predictive accuracy.
- A small value of ϵ will result in a complex model that closely fits the training data, but risks overfitting the training set and therefore generalizing poorly to the new data.
- On the other hand, a large value of ϵ will result in a simpler model with larger errors but potentially a better generalization performance.
- To identify a robust choice of ϵ, a common approach is cross-validation, which involves partitioning the raw data into training and validation sets several times, each starting with a different random seed.

Lab session: implementing pairs trading using rules and machine learning models