Pisco Med Publishing

Application of Fama-Fench three-factor model in Chinese A-share market--Based on SVM machine learning model

Peng Yun Wang Guan Ying Dai Wei Xin Zhang

(NanKai University TianJin 300071)

448900879@qq.com

Abstract: How to apply machine learning in the field of financial investment has been a hot research topic in academia and finance. In this paper, the support vector machine method (SVM) in machine learning is combined with Fama-Fench three-factor model to construct a new quantitative investment strategy, and the empirical analysis is carried out by using A-shares. Research shows that support vector machine (SVM) combined with the traditional three-factor model can build a more effective portfolio.

Keywords: Machine Learning; Fama-Fench three-factor model; SVM

1. Introduction to Model Theory

1.1 Fama-Fench three-factor model

The capital asset pricing model (CAPM) proposed by Sharpe (1964), Lintner (1965) and Mossin (1966) is a milestone. Under certain assumptions, they rigorously deduced the pricing formula of any securities in the equilibrium state:

$$E(r_i) = E(r_0) + \beta_i[E(r_m) - E(r_0)]$$

Fama (1973) verified CAPM and found that the β value of the combination and its return rate's linear relationship is approximately true, but the intercept is high and the slope is low, indicating that β cannot be solved[1]. Then Fama & Fench (1992) analyzed the causes in detail about the influence of CAPM anomaly factor on cross-sectional return of securities. The results show that all these factors have a separate explanatory power on the cross-sectional return rate, but when combined, the market value and accounts for the objective-to-value ratio (BE/ME), two factors largely absorb the estimated ratio (E/P) and the role of the leverage ratio. Based on this, Fama & Fench (1993) built a multi-factor model [2]. The two factors of SMB and HML are considered. For here, the three-factor model can be written as:

$$E(R_i) - R_f = b_i\left[E(R_m) - R_f\right] + s_i E(SMB) + h_i E(HML)$$

1.2 Support Vector Machines (SVM)

Support vector machine (SVM) is the most successful machine learning method in the 1990s. Its basic idea is to solve the separating hyperplane which can correctly divide the training data set and has the largest geometric interval. The hyperplane can be used for data. The standard of classification originates from logistic regression, which aims at feature learning 0/1 classification model, logistic function (sigmoid function) can be expressed as[3] :

$$f(x) = \sigma(wx + b) = \frac{1}{1 + e^{-(wx+b)}}$$

This model takes the linear combination of features as independent variables. Due to the value of independent variables The range is from negative to positive infinity, so the sigmoid function maps independent variables to (0,1), the corresponding category is represented by y, which can be taken as −1 or 1. Logarithm based on probability of loss according to the classification, and the sigmoid function image is shown in Figure 1.

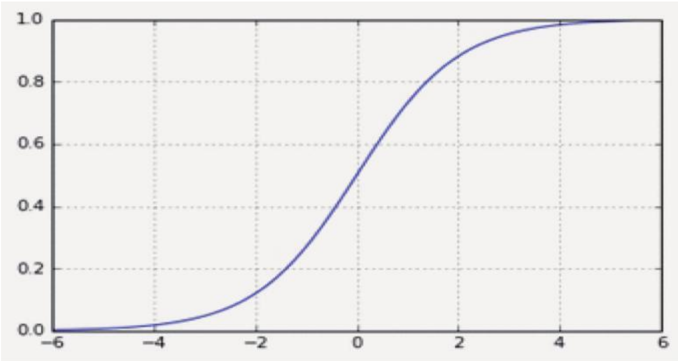

Figure 1

Support Vector Machine also uses the above classification principle to classify data. Constructing a maximum margin classifier γ to maximize the distance between two margin boundaries and fall on the margin. The point is called support vector, obviously $y(w^T + b) > 1$

When the data cannot be linearly separable, it is necessary to use the nonlinear model to classify the data well. When the data cannot be separated by a straight line, constructing a hypersurface can separate the data. SVM adopts the method of selecting a kernel function, by mapping the data to high-dimensional space, constructing the optimal classification hyperplane in this space, using linear classification method for data classification.

However, it is difficult to determine which kernel function is appropriate without knowing the form of feature mapping. Therefore, the selection of different kernel functions may face different results. If the selection of kernel functions is not appropriate, it means that the sample is mapped to an inappropriate feature space, which may lead to poor results. Common kernel functions are shown in Table 1

Linear kernel	$k(x_i,x_j) = x_i^T x_j$
Polynomial kernel	$k(x_i,x_j) = (x_i^T x_j)^d$
Gaussian amplifier kernel	$k(x_i,x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$
Laplacian kernels	$k(x_i,x_j) = \exp\left(-\frac{\|x_i - x_j\|}{\sigma}\right)$
Sigmoid kernels	$k(x_i,x_j) = \tanh(\beta x_i^T x_j + \theta)$

2. Empirical Analysis and Application

2.1 Data declaration

The training data are from the stock factor values of Shanghai and Shenzhen 300 constituent stocks traded on the last trading day of each month from August 1, 2012 to August 1, 2016. The market value factor SMB and book value ratio are shown in Table 2. All factor data are standardized and processed. Using PB and market Value two factors to predict the stock ' s price and price next month, using support vector machine in machine learning to train. The rise and fall of each trading day are 1, and the decline and stock price are marked as 0. The cross-validation method is adopted, in which 80% of the data are the training set and 20 % of the data are the test set. The e1071 package in R is used for analysis.

Factor number	Factor name	Factor type	Factor description	
1	marketValue	Scale factor	Market value of logarithmic circulation	TTM algorithm
32	PB	Valuation factor	Net market rate	TTM algorithm

2.2 Strategy retest

The experimental results show that the prediction accuracy of SVM test set is 62.32 %, and the back-test strategy is to buy the top 20 stocks with equal weight in the month 's forecast rise probability, and adjust the warehouse on the first trading day of the early month. The back-test interval is from July 1, 2013 to February 28, 2017, and the initial capital is set to 1000000 yuan. The back-test is carried out by using the quantitative platform of excellent ore. Table 2 is partial hold records for policy backtesting .Policy effects are shown in Figure 2 and Table 3.

Table1:

Date	Code	Number of warehouses held	Mean price	Closing price	Cash position
2013/7/1	61	10400	5.88	5.93	69933.25
2013/7/1	2385	8300	4.45	4.51	37441.3
2013/7/1	60058	58100	6.01	6.25	363125
2013/7/1	601888	600	28.26	29.37	17619
2013/7/1	6001888	17000	9.27	9.44	160548
...	...	...	...	...	...
2017/2/28	600754	2400	32.7	31.07	74568
2017/2/28	2299	2900	29.48	18.59	53911
2017/2/28	300182	11100	11.5	10.2	113220
2017/2/28	600304	10500	25.32	25.62	269010
2017/2/28	600252	68100	4.35	4.73	322113

Figure 2:

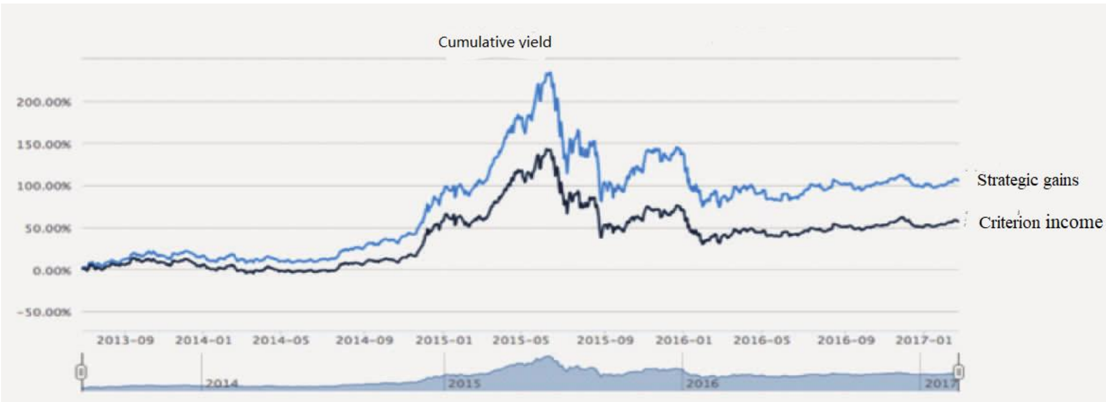

Table 3：

Annualized return rate	Benchmark annualized yields	Alpha	Beta	Sharpe ratio	Yield volatility	Information ratio	Maximum retracement	Buying-selling rate
22.4%	13.4%	8.2%	1.07	0.64	29.4%	1.29	48.1%	8.94

Since the multi-factor model is usually a robust strategy, in order to avoid the high transaction costs caused by frequent transactions, this strategy adopts the monthly regular position adjustment. According to the retest results of the strategy in Figure 2, the transaction strategy designed by the support vector machine algorithm combined with the Fama-Fench three-factor model has an annualized return rate of 22.4 % in the retest interval, which exceeds the benchmark market return rate of 13.4 % and obtains the Alpha of 8.2 %, indicating that the Fama-Fench three-factor model is still effective in the A-share market. At the same time, we can see that the maximum withdrawal of the strategy is 48.1 %, indicating that the strategy cannot achieve good hedging effect without adding stop loss and stop surplus. From the perspective of quantitative investment, hedging with stock index futures is a good choice for multi-factor strategy.

3. Conclusion

The research shows that machine learning method has a good application space in the financial market. In the era of big data, the traditional statistical model cannot extract effective information features from complex and multidimensional financial data, and machine learning algorithm is good at dealing with complex and high dimensional data. This is also the reason why artificial intelligence investment has more and more attention in the financial industry. From the perspective of quantitative investment, how to apply machine learning method to the field of financial investment is still a controversial topic, this paper only from the perspective of trying, innovative machine learning method combined with the classic Fama-Fench three-factor model to verify the investment effect of Chinese stock market.

Reference:

[1] Fama ,E.F.& K.R .French(1992), "The cross-section of expected stock returns", Journal of Finance 47:427-466.

[2] Fama , E.F.& K.R .French (1993), "Common risk factors in the returns on stocks and bonds" , Journal of Financial Economics 3:33-56.

[3] Cortes C, Vapnik V. Support-Vector Networks.[J].Machine Learning，1995，20(3):273-297.