Stock Market Prediction Using Dynamic Barycenter Averaging Kernel in RBF Networks

M.S. Student: Yu-Ting Lan(藍語庭), Department of Applied Mathematics

摘要

在動盪的金融市場中，預測股票市場的未來趨勢是很重要的。在本篇論文中，提出將動態重心平均核(DBAK)引入徑向基底函數網絡(RBFN)。首先，我們將 K-means 和 DTW 重心平均(DBA)相結合，以確定 DBAK 的中心。然後，我們使用基於 DBAK 的徑向基底函數網絡，並利用四個徑向基底函數(RBFs)做預測，分別是高斯、多元二次曲面、逆二次、逆多元二次曲面。最後，預測結果顯示在四個徑向基底函數中逆二次函數預測效果最佳。

Abstract

In a volatile financial market, it is important to predict the future trend of the stock market. In this paper, we propose to introduce the dynamic barycenter averaging kernel (DBAK) into the Radial Basis Function Network (RBFN). First, we combine K-means and DTW barycenter averaging (DBA) to determine the center of DBAK. Then, we use the DBAK based on Radial Basis Function Network and make predictions using four Radial Basis Functions (RBFs), which are Gaussian, multi-quadratic, inverse quadratic, and inverse multi-quadratic. Finally, the prediction results show that the inverse quadratic function has the best prediction among the four radial basis functions.

Data sources

We use the data from the yahoo finance website. It is including open, high, low, close, adjusted close, volume values. We chose two companies, Apple Inc. (AAPL) and Microsoft Inc. (MSFT), the period ranges from 2020/1/2 to 2020/12/30. The following charts show the closing price trend of the two companies, the *x*-axis is the total 252 days, the *y*-axis is the closing price.

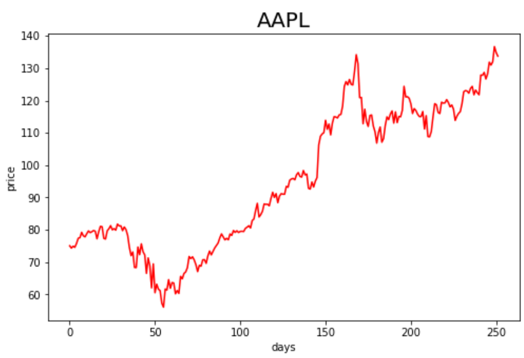

Fig. 1. The price trend for AAPL.

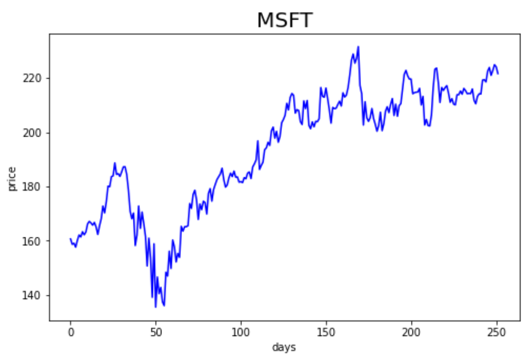

Fig. 2. The price trend for MSFT.

Method

We use the DBAK-RBFN model which was developed by Kejian Shi [1]. We combine K-means [2] and DBA [3] to determine the center of DBAK. We cluster the given data x by K-means and then using the DBA algorithm to obtain the medoid of time-series in each cluster.

$$\underset{S}{arg\ min} \sum_{i=1}^{k} \sum_{x \epsilon S_i} \|x - \mu_i\|^2$$

where x is the observation data, $S = \{S_1, S_2, \dots, S_k\}$ is a set of the k cluster, μ_i is the mean of points in S_i, k is the cluster.

And then, we use the DBAK-RBFN model, which including four RBFs(φ) to predict the close price: Gaussian (G), Multiquadric (MQ), Inverse quadratic (IQ), Inverse multiquadric (IMQ). The performance of the model we determine by mean absolute error (MAE).

$$\varphi^{\mathrm{G}} = e^{\frac{-(x-c)^2}{2\gamma^2}} \qquad \varphi^{\mathrm{MQ}} = \sqrt{1 + \gamma^2 (x - c)^2}$$

$$\varphi^{\mathrm{IQ}} = \frac{1}{1+\gamma^2(x - c)^2} \qquad \varphi^{\mathrm{IMQ}} = \frac{1}{\sqrt{1+\gamma^2(x - c)^2}}$$

where γ is the width of distribution, c is centers, w_j is the j^{th} linear weight, and $\hat{y} = \sum_{j=1}^{k} w_j \varphi_j$, $\hat{y}$ is the predicted price.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \|\hat{y}_i - y_i\|$$

where $\hat{y}_i$ is the i^{th} predicted value, and y_i is the i^{th} real market price.

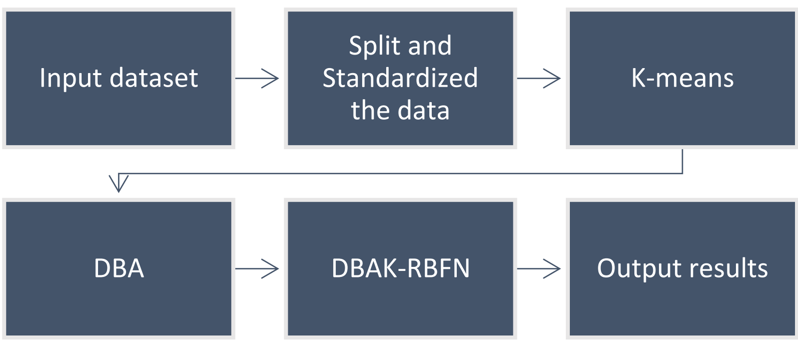

Fig. 3. The general structure of RBF Network with DBAK.

Experimental results

We split the time series data into a 70% training set and a 30% test set and use normalization into the training set and the test set. And then, we use K-means and the DBA algorithm.

From AAPL Inc. data, in Fig. 4., when k=7 can get the highest silhouette coefficient (SC). In Fig. 5., we use the DBAK in each cluster, it represents the set of time series from the training set that was assigned to the considered cluster (in black) as well as the barycenter of the cluster (in red), the x-axis is the time-series(ts), the y-axis is the signal value(sv). Finally, we select the RBFs with the smallest MAE values. In Tab. 1. and Fig. 6., the IQ is the lowest MAE value.

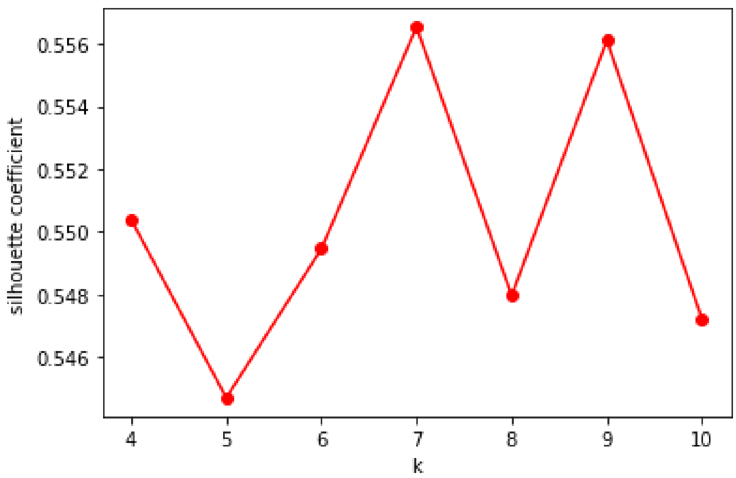

Fig. 4. Use SC to determine k for AAPL (k = 7).

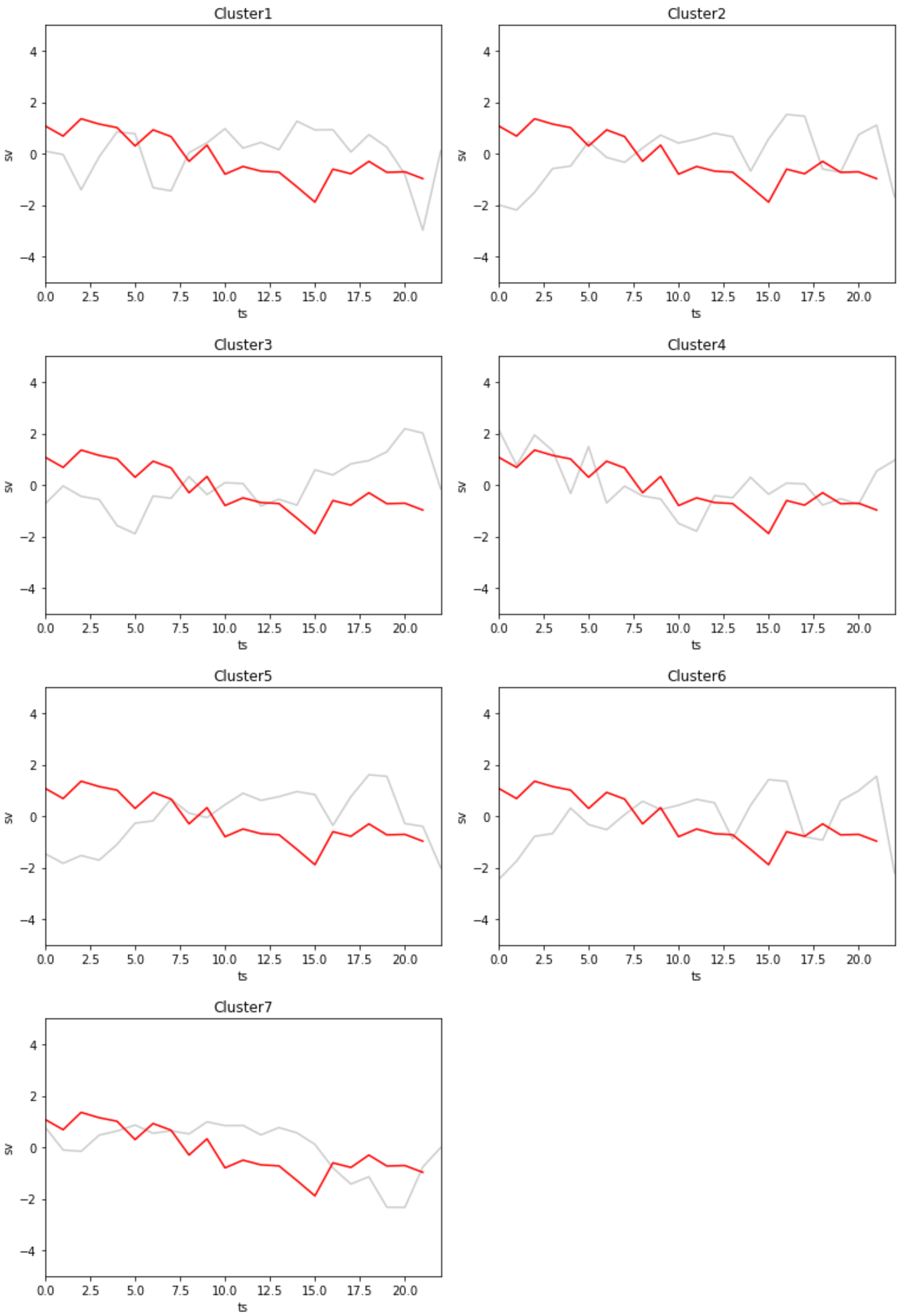

Fig. 5. Use DBA with K-means (*k* = 7) to determine the center for AAPL.

	G	MQ	IQ	IMQ
MAE	0.0825	0.0569	0.0495	0.0658

Tab. 1. The MAE of DBAK-RBFN with various RBFs for AAPL.

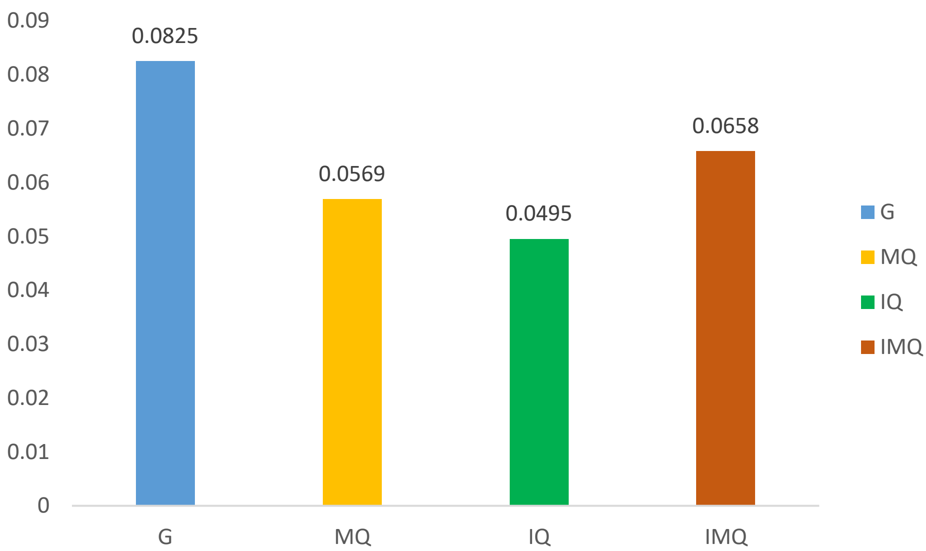

Fig. 6. The MAE of DBAK-RBFN with various RBFs for AAPL.

From MSFT Inc. data, in Fig. 7., when k=5 can get the highest SC. In Fig. 8., we use the DBAK in each cluster. Finally, we select the RBFs with the smallest MAE values. In Tab. 2. and Fig. 9., IQ is the lowest MAE value.

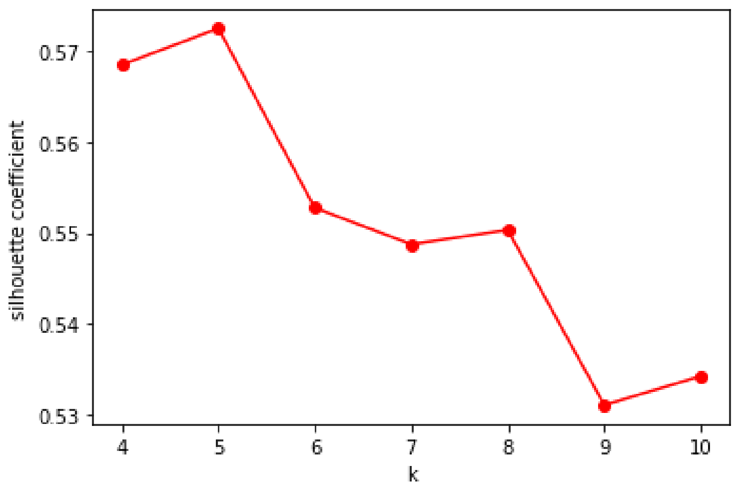

Fig. 7. Use SC to determine *k* for MSFT (*k* = 5).

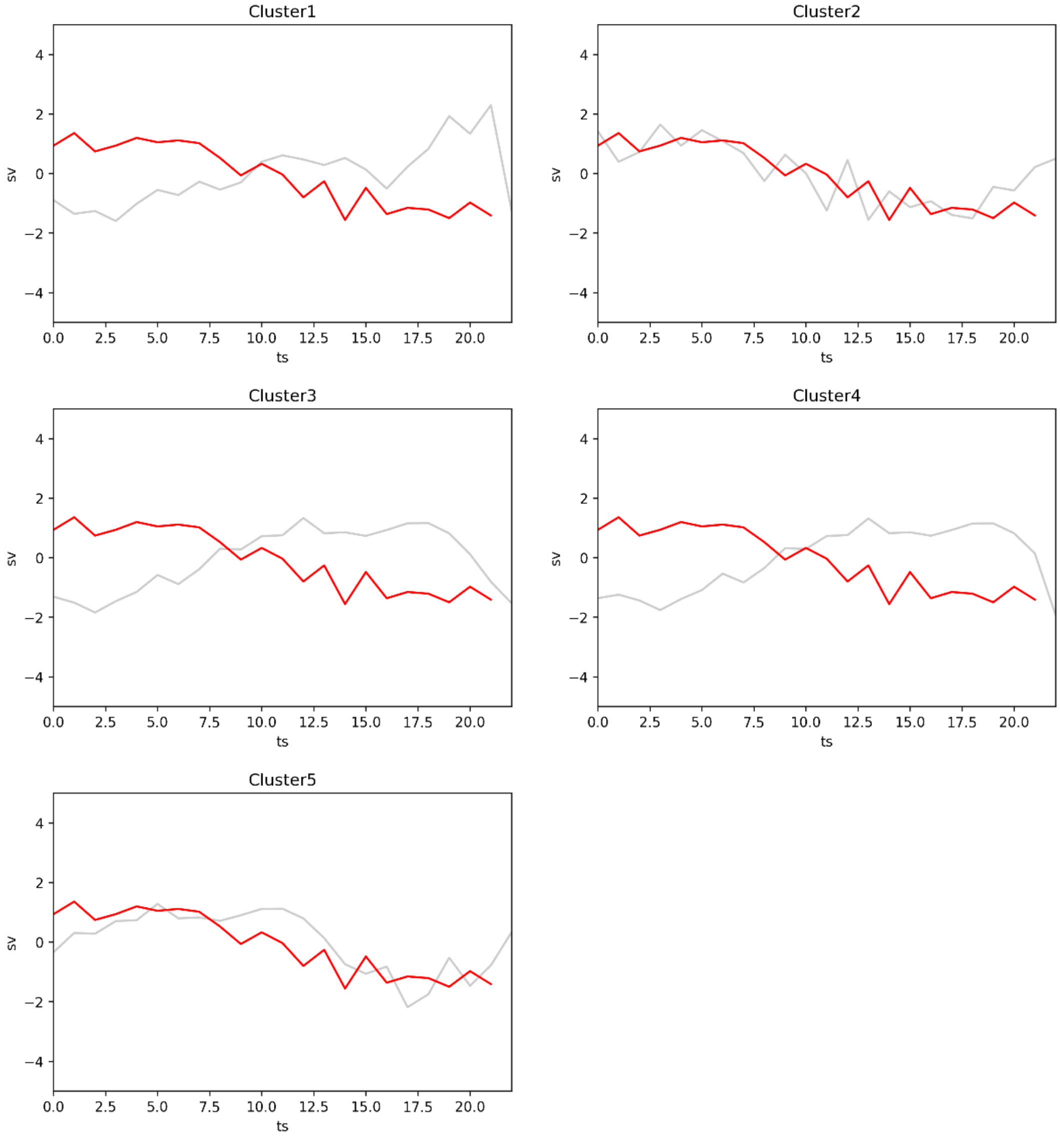

Fig. 8. Use DBA with K-means (*k* = 5) to determine the center for MSFT.

	G	MQ	IQ	IMQ
MAE	0.0658	0.0621	0.045	0.059

Tab. 2. The MAE of DBAK-RBFN with various RBFs for MSFT.

教授簽名:

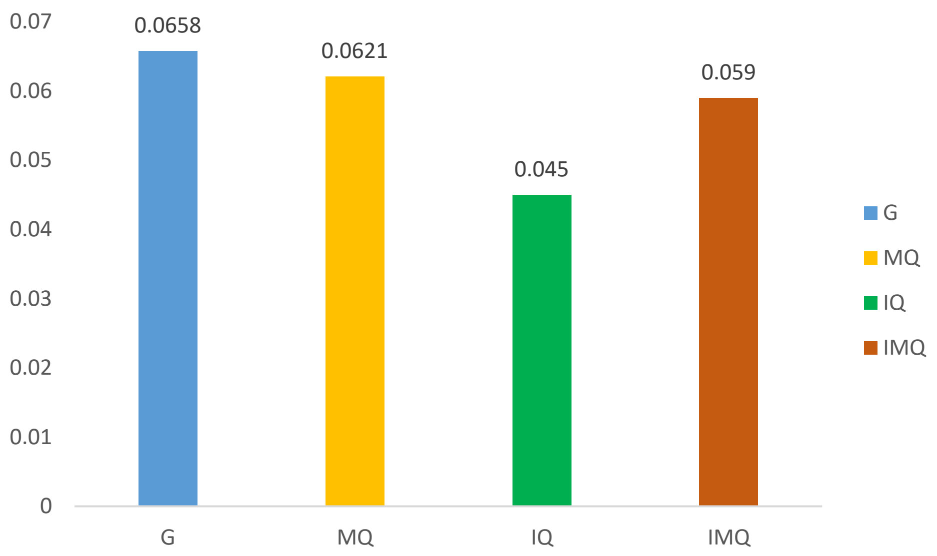

Fig. 9. The MAE of DBAK-RBFN with various RBFs for MSFT.

Conclusions

In this paper, we use the DBAK-RBFN and make predictions using four RBFs. When *k*=7 from the data of Apple Inc. and *k*=5 from the data of MSFT Inc., we can get IQ values of 0.0495 and 0.045 respectively. Finally, the prediction results show that the inverse quadratic function has the best prediction among the four radial basis functions.

References

[1] Kejian Shi, Dynamic Barycenter Averaging Kernel in RBF Networks for Time Series, vol. 7, pp. 47564 – 47576, 2019

[2] Peter T. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, Journal of Computational and Applied Mathematics, vol. 20, pp. 53-65, 1987.

[3] D. J. Berndt, “Using dynamic time warping to find patterns in time series,” in Kdd Workshop, 1994, pp. 359–370.

[4] D.S. Broomhead and D. Lowe, “Multivariable Functional Interpolation and Adaptive Networks ", Journal of Complex Systems, vol. 2, pp. 321-355, 1988.

[5] J. A. Hartigan and M. A. Wong. " Algorithm AS 136: A K-Means Clustering Algorithm ", Journal of the Royal Statistical Society. Series C (Applied Statistics), vol. 28, pp. 100–108, 1979.

教授簽名: