Original Article
International Journal of Fuzzy Logic and Intelligent Systems
Vol. 16, No. 1, March 2016, pp. 44-51
http://dx.doi.org/10.5391/IJFIS.2016.16.1.44

ISSN(Print) 1598-2645
ISSN(Online) 2093-744X

Some Observations for Portfolio Management Applications of Modern Machine Learning Methods

Jooyoung Park[1], Seongman Heo[1], Taehwan Kim[1], Jeongho Park[1], Jaein Kim[2], and Kyungwook Park[3]
[1]Department of Control & Instrumentation Engineering, Korea University, Sejong City 339-700, Korea
[2]Department of Mathematics, Korea University, Seoul 136-713, Korea
[3]School of Business Administration, Korea University, Sejong City 339-700, Korea

Abstract

Recently, artificial intelligence has reached the level of top information technologies that will have significant influence over many aspects of our future lifestyles. In particular, in the fields of machine learning technologies for classification and decision-making, there have been a lot of research efforts for solving estimation and control problems that appear in the various kinds of portfolio management problems via data-driven approaches. Note that these modern data-driven approaches, which try to find solutions to the problems based on relevant empirical data rather than mathematical analyses, are useful particularly in practical application domains. In this paper, we consider some applications of modern data-driven machine learning methods for portfolio management problems. More precisely, we apply a simplified version of the sparse Gaussian process (GP) classification method for classifying users' sensitivity with respect to financial risk, and then present two portfolio management issues in which the GP application results can be useful. Experimental results show that the GP applications work well in handling simulated data sets.

Keywords: Machine learning, Gaussian processes, Portfolio management

Received: Mar. 1, 2016
Revised : Mar. 15, 2016
Accepted: Mar. 24, 2016

Correspondence to: Jooyoung Park
(parkj@korea.ac.kr)

1. Introduction

Recently, artificial intelligence has reached the level of top information technologies that will have significant influence over many aspects of our future lifestyles. In particular, in the fields of machine learning technologies for classification and decision-making, there have been a lot of research efforts for solving estimation and control problems that appear in the various kinds of portfolio management problems via data-driven approaches. Note that these modern data-driven arpproaches, which try to find solutions to the problems based on relevant empirical data rather than mathematical analyses, are useful particularly in practical application domains.

In this paper, we consider the problem of applying kernel methods together with some other optimization methods for portfolio management. As well-known, kernel methods have attracted great interests in the areas of pattern classification, function approximation, and anomaly detection [1–9], and recently Gaussian process played an important role in the field of machine learning as a tool for probabilistic kernel methods [10]. We apply a simplified version of the sparse Gaussian process (GP) classification method, which is a direct result of two recent remarkable Gaussian process papers [25, 26], for performing risk sensitivity classification in dealing with financial portfolio management. Since portfolio management problems are optimal decision-making problems that rely on actual empirical data, theoretical and practical solutions can be formulated via many of recent machine learning and control advancements: the traditional mean-variance efficient portfolio problem [11]; index tracking portfolio formulation [12–15]; risk-adjusted expected return maximizing strategy [16–18]; trend following strategy [19–23]; long-short trading strategy (including the pairs trading strategy) [20, 24], etc. In this paper, we also raise two important portfolio management issues in which the GP application results can be useful.

This paper is organized as follows: In Section 2, we briefly describe relevant GP preliminaries. Applying a simplified version of the sparse Gaussian process (GP) classification method for performing risk sensitivity classification as well as their possible applications to portfolio management issues are presented in Section 3. Finally, in Section 4, we present our concluding remarks.

2. Preliminaries

Probabilistic kernel methods, which include Gaussian processes, have recently attracted great interests in the areas of pattern classification, function approximation, and anomaly detection. In this section, we briefly describe some preliminaries on Gaussian processes, which plays an important role in our portfolio management applications. For more details on the Gaussian processes, please refer to, e.g., [10]. Gaussian process, $\{f(x)\}$, is an indexed family of random variables with index $x \in R^d$ such that for any finite indices, $x_1, \cdots, x_N$, $f(x_1), \cdots, f(x_N)$ are jointly Gaussian. Gaussian processes can be characterized by their mean functions and covariance (or kernel) functions, which are defined as follows, respectively:

$$m(x) = E[f(x)], \tag{1}$$

$$k(x, x') = E[(f(x) - m(x))(f(x') - m(x'))]. \tag{2}$$

Gaussian processes with mean function $m(x)$ and covariance function $k(x, x')$ are often denoted by

$$f(x) \sim GP(m(x), k(x, x')). \tag{3}$$

One can see that with the so-called kernel trick, Bayesian linear models defined on the feature space can be viewed as Gaussian processes. More specifically, let's suppose that $f(x)$ is described by $\phi(x)^T w$, where the prior distribution of the random vector w is $N(0, \Sigma_w)$. Here, $\phi(x)$ is the feature vector, which is the result of mapping the input vector x into the (possiably high-dimensional) feature space F. Note that in this situation, the expectation of $f(x)$ is

$$E[f(x)] = E[\phi(x)^T w] = \phi(x)^T E[w] = 0, \tag{4}$$

and the covariance between $f(x)$ and $f(x')$ satisfies

$$\begin{aligned} E[f(x)f(x')] &= E[\phi(x)^T w w^T \phi(x')] \\ &= \phi(x)^T E[ww^T]\phi(x') = \phi(x)^T \Sigma_w \phi(x'). \end{aligned} \tag{5}$$

Thus defining the kernel function, k, by the kernel trick

$$k(x, x') = \phi(x)\Sigma_w \phi(x') \tag{6}$$

enables us to compute the covariance, $\mathrm{Cov}[f(x), f(x')]$, directly on the input space using the kernel, i.e., $\mathrm{Cov}[f(x), f(x')] = k(x, x')$. Therefore, k can be conveniently interpreted as both a kernel function (in the sense of kernel methods) and a covariance function. In general, the mean function, $m(x)$, is assumed to be the zero function, and the assumption is thought to be without loss of generality. The task of obtaining the predictive distribution for any test point in the Gaussian process framework can be summarized as follows: Consider the training data set $D = \{(x_n, y_n)\}_{n=1}^N$, where $X = \{x_n \in R^d\}_{n=1}^N$ is the set of the input values of the training data, and $y = \{y_n \in R^d\}_{n=1}^N$ is the set of the corresponding target values. Since the Gaussian process $f(x)$ has the zero mean function and the kernel function, $k(x, x')$, the joint distribution of the random vector $f = [f(x_1), \cdots, f(x_N)]^T$ can be written as

$$p(f|X) = N(f|0, K(X)). \tag{7}$$

Here by $N(f|m, V)$, we mean the multi-variate Gaussian distribution with mean vector m and covariance matrix V. Also, $K(X)$ is an $N \times N$ matrix, whose (i, j)-th element is $k(x_i, x_j)$. For notational convenience, we often use K instead of $K(X)$. In this paper, we consider the following squared exponential (SE) kernel, which is one of the most widely used choices in the kernel method community:

$$k(x_i, x_j) = \sigma_f^2 \exp[-\frac{1}{2l^2}(x_i - x_j)^T(x_i - x_j)]. \tag{8}$$

Here σ_f and l, which charaterize the shape of the kernel function, are called hyper-parameters, and the vector consisting of hyper-parameters is denoted as θ. In the Gaussian process regression, the disturbance which occurs in the process of the data observation is taken into account too, and it is characterized by means of a Gasussian noise model:

$$p(y|f) = N(y|f, \sigma_n^2 I). \tag{9}$$

Hence, by combining $p(f|X)$ and $p(y|f)$, one can obtain the following marginal likelihood for the regression problem: $p(y|X) = N(y|0, K + \sigma_n^2 I)$. Also, the log marginal likelihood for the whole training data D can be written as follows:

$$\log p(y|X) = -\frac{1}{2} y^T (K + \sigma_n^2 I)^{-1} y \\ -\frac{1}{2}\log|K + \sigma_n^2 I| - \frac{N}{2}\log(2\pi). \tag{10}$$

Finding the optimal hyper-parameter vector can be achieved by maximizing the above log marginal likelihood function with respect to θ. Also, the predictive distribution of the ouput y_* for the test input point x_* can be obtained by applying the conditional density formula for the multi-variate Gaussian distributions [10], i.e.,

$$p(y_*|x_*, D) = N(y_*|k_*^T (K + \sigma_n^2 I)^{-1} y, \\ k_{**} - k_*^T (K + \sigma_n^2 I)^{-1} k_* + \sigma_n^2). \tag{11}$$

Here, k_* and k_{**} are used for notational convenience, and they mean the following, respectively:

$$k_* = [k(x_1, x_*), \cdots, k(x_n, x_*)]^T, \tag{12}$$

$$k_{**} = k(x_*, x_*). \tag{13}$$

Finally, note that the point esimate $k_*^T (K + \sigma_n^2 I)^{-1} y$, which is the mean of y_*, can be further written as

$$\hat{y}_* = \sum_{i=1}^{N} \alpha_i k(x_i, x_*), \tag{14}$$

where $\alpha = [\alpha_1, \cdots, \alpha_N]^T = (K + \sigma_n^2 I)^{-1} y$, and that (14) can be viewed as a result of the representer theorem [1–3] of the kernel methods.

3. Applications

In this section, we present some observations for portfolio management applications of Gaussian processes, natural evolution strategy, and Hamilton-Jacobi-Bellman (HJB) equations. Our observations consist of two parts. In the first part, we consider the applicability of a simplified sparse Gaussian process classification (GPC) method, which is a direct result of two recent remarkable Gaussian process papers [25, 26], for the task of classifying individuals' sensitivity with respect to financial risk. Derivation of the simplified sparse GPC method can be summarized as follows: We consider the input data set $X = \{x_n\}_{n=1}^N$ together with the target data set $Y = \{y_n\}_{n=1}^N$, where $x_n \in R^d$ and $y_n \in \{1, \cdots, C\}$. Note that the n-th observation, y_n, is a categorical variable that can be transformed into the one-hot-encoding format. Also, note that for the observation y_n, one can use a multinomial distribution whose probabilities are defined by softmax having intensities $f_n = (f_{n1}, \cdots, f_{nC})$. The k-th intensity of f_n, f_{nk}, is the ouput of the Gaussian process $F_k(x_n)$. To achieve a sparse representation, the so-called inducing points, $Z \in R^{M \times d}$, are introduced. Note that in classification problems, the marginal log-likelihood is not tractable, contrary to the case of Gaussian process regression of Section 2. Hence, we need to rely on a variational approximation. With $q(f, U) = q(U)p(f|X, U)$ and Jensen's inequality [10], we have

$$\log p(Y) = \log \int p(U)p(f|X,U)p(Y|f) df dU \\ \geq \int q(U)p(f|X,U) \log \frac{p(U)p(f|X,U)p(Y|f)}{q(U)p(f|X,U)} df dU \\ = -KL[q(U) \parallel p(U)] \\ + \sum_{n=1}^{N} \int q(U)p(f_n|x_n, U) \log p(y_n|f_n) df_n dU, \tag{15}$$

where KL stands for Kullback-Leibler divergence [10]. Note that with $q(U) = N(U|m, S)$, we have

$$p(f_n|x_n, U) = \prod_{k=1}^{C} N(f_{nk}|a_n^T u_k, b_n), \tag{16}$$

where $a_n = K_{MM}^{-1} K_{Mn}$, $b_n = K_{nn} - K_{nM} K_{MM}^{-1} K_{Mn}$. In this paper, we consider the class of diagonal covariance matrices for S and called the resultant GPC a simplified sparse Gaussian process classification. Since the integration of $\log p(y_n|f_n)$ in the right hand side of (15) is not tractable, we rely on the sampling-based approximation of [27, 28]. In this paper, we propose to use the simplified sparse GPC as a framework of classifying users' sensitivity with respect to financial risk. The categorical target variable in the framework describes the sensitivity level (e.g., very sensitive to risk, moderately sensitive to risk, only a little sensitive to risk, etc). The questions for providing inputs along the line may include the following kinds [32–35]:

1. What is your current age and planned age of retirement?
2. Your annual before-tax income is $ _______.
3. Your future income until your retirement will be ____.
   (a) Increasing
   (b) The same
   (c) Decreasing
   (d) Unpredictable
4. Total value of your cash and other liquid securities is $ _______.
5. Your investment horizon is _______ years.
6. Your primary investment objective is _____.
   (a) Investing for comfortable retirement
   (b) General investing for wealth accumulation
   (c) Securing an emergency fund
   (d) Saving for a specific purpose (for example college education for kids)
7. Your tolerance for risk taking when investing is _____.
   (a) Defensive - You accept lower returns to protect your initial investment.
   (b) Moderate - You want balance between the stability and long-term return.
   (c) Assertive - You are prepared to accept higher volatility to accumulate assets over long term.
8. If your entire investment portfolio lost 10% of its value in a month during a market decline, what would you do?
   (a) Liquidate all the investment
   (b) Sell half of the portfolio
   (c) Keep the portfolio
   (d) Invest more
9. What return do you expect to achieve from your investments?
   (a) Return without losing original money
   (b) 3-6% per annum
   (c) 7-10% per annum
   (d) 11-15% per annum
   (e) Over 15% per annum

Figure 1. Training data considered for binary classification.

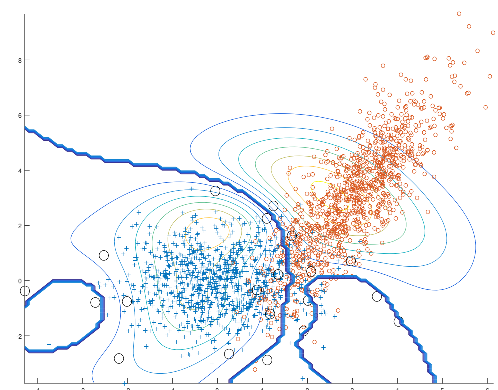

Figure 2. Classification results of the simplified sparse GPC (with 20 inducing points).

In order to evaluate the validity and strengths of the simlified sparse GPC method, we performed experiments for the simulated data (see Figs. 1-9). From the classification results, one can see that the simplified sparse GPC work well with relatively small number of inducing points. Also, Figs. 2-5 show that the GPC can achieve sparsity somewhat more efficiently compared to the standard SVM approach (for which we used *fitcsvm* of MATLAB).

In the second part of our observations, we present two portfolio management issues that can utilize the simplified sparse GPC method. The issues covered along the line are the trend-following problem [19, 20, 29] and the portfolio optimization problem [30]. In the first issue, we consider an exponential natural evolution strategy (NES) [31] based solution to find

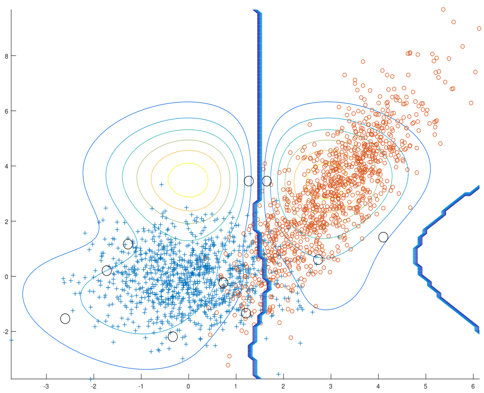

Figure 3. Classification results of the simplified sparse GPC (with 10 inducing points).

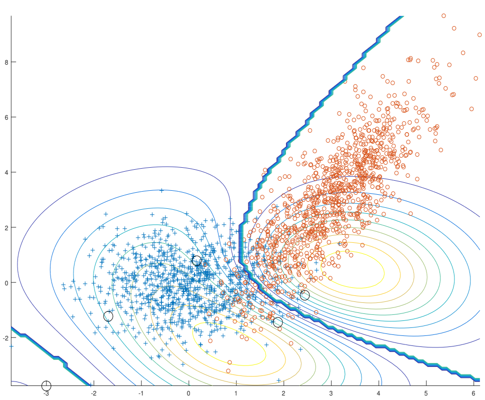

Figure 4. Classification results of the simplified sparse GPC (with 5 inducing points).

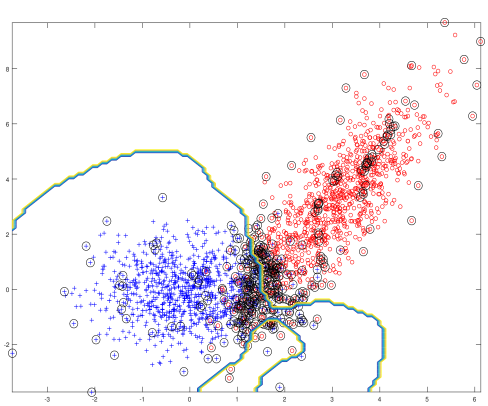

Figure 5. Classification results of SVM (with 392 support vectors).

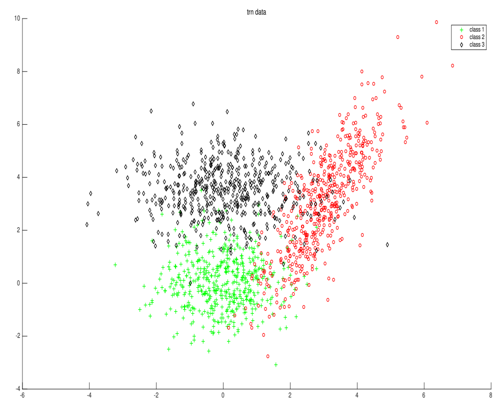

Figure 6. Training data considered for multi-class classification.

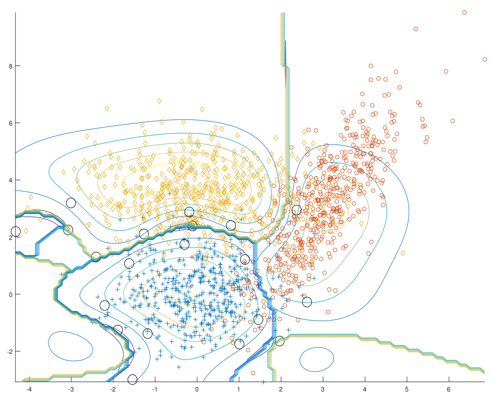

Figure 7. Classification results of the simplified sparse GPC (with 20 inducing points).

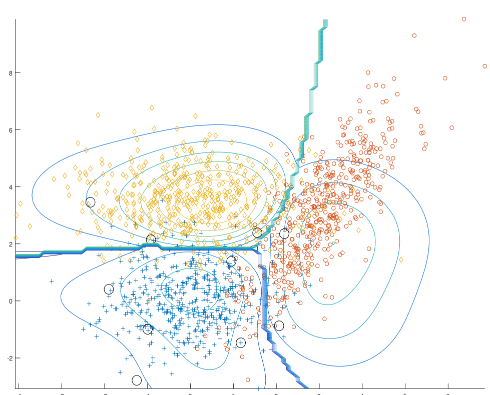

Figure 8. Classification results of the simplified sparse GPC (with 10 inducing points).

an efficent trend following strategy (For details, please refer to [20, 29]), and propose the strategy of using the transaction cost, K, as a tuning parameter that can vary according to the GPC results (Fig. 10). In the second issue, we consider a HJB equation based portfolio optimization problem [30], where a user has the choice of investing in the stock market or saving

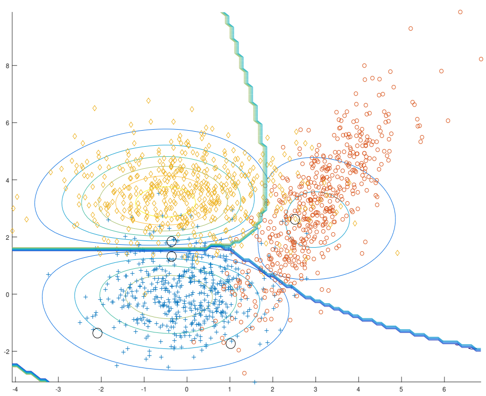

Figure 9. Classification results of the simplified sparse GPC (with 5 inducing points).

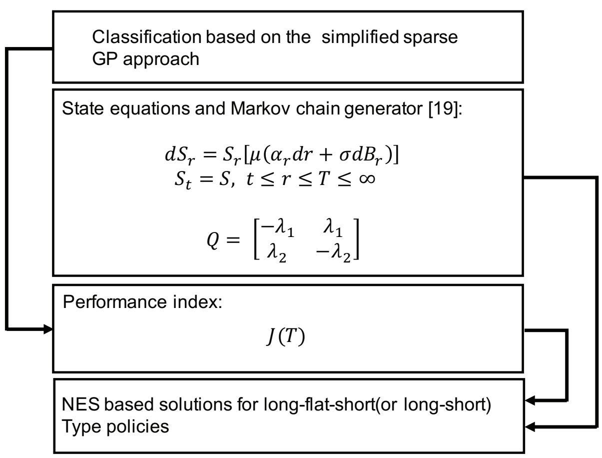

Figure 10. Conceptual diagram of the first issue.

in a bank account, the stock market is modelled as geometric Brownian motion, and dynamics for the factor and the volatility are also modelled with appropriate stochastic differencial equations (For details, please refer to [30]), and propose the strategy of using the coefficient of risk aversion, γ, as a tuning parameter that can vary according to the GPC results (Fig. 11). We expect that in our future works, these two issues will ultimately lead to a set of fundamental building blocks for efficient personal financial planning packages.

4. Conclusion

Modern data-driven machine learning approaches, which try to find solutions to the problems based on relevant empirical data rather than mathematical analyses, are useful particularly in practical application domains. In this paper, we apply a simplified version of the sparse Gaussian process (GP) classification method to two portfolio management issues (NES-based trend-following, HJB-based porfolio optimization). Experimental results showed the applicability of the simplified sparse GPC in simulated data sets. For future works, we are planning to consider more extensive simulation studies, which will show the strengths and weaknesses of the proposed idea, and applications of our methods to an integral package that can deal with personal financial planning problems.

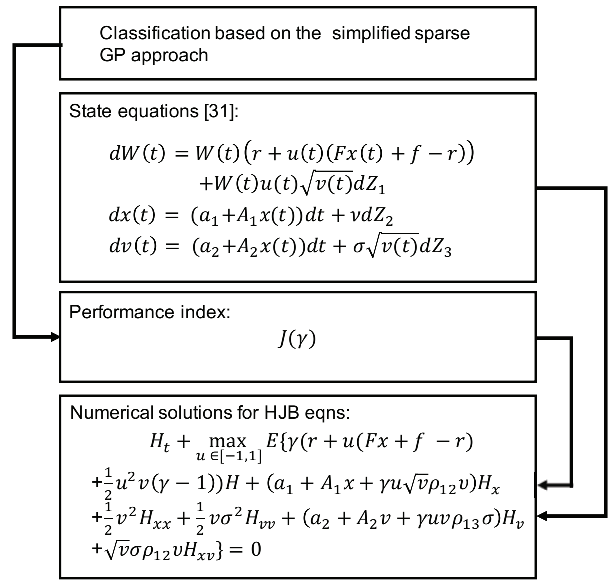

Figure 11. Conceptual diagram of the second issue.

Acknowledgement

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0021188).

References

[1] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.

[2] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, 2000.

[3] B. Schőlkopf and A. J. Smola, *Learning with Kernels*, MIT Press, 2002.

[4] K.-R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Schlkopf, “An introduction to kernel-based learning algo-

rithms," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 181-201, 2001.

[5] J. T. Kwok, "The evidence framework applied to support vector machines," *IEEE Transactions on Neural Networks*, vol. 11, no. 5, pp. 1162-1173, 2000.

[6] B. Schőlkopf, J. C. Platt, J. Shawe-Taylor, and A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, pp. 1443-1471, 2001.

[7] B. Schőlkopf, A. Smola, R. Williamson, and P. L. Bartlett. "New support vector algorithms," *Neural Computation*, vol. 12, no. 5, pp. 1207-1245, 2000.

[8] J. Park, J. Kim, H. Lee, D. Park, "One-class support vector learning and linear matrix inequalities," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 3, no. 1, pp. 100-104, June 2003.

[9] J. Park and D. Kang, "A Modified approach to density-induced support vector data description," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 7, no. 1, pp. 1-6, March 2007.

[10] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, 2006.

[11] H. M. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*, Cowles Foundation Monograph, No. 16, John Wiley andSons, 1959.

[12] J. A. Primbs, and C. H. Sung, "A stochastic receding horizon control approach to constrained index tracking," *Asia-Pacific Financial Markets*, vol. 15, pp. 3-24, 2008.

[13] J. E. Beasley, N. Meade, and T.-J. Chang, "An evolutionary heuristic for the index tracking problem," *European Journal of Operational Research*, vol. 148, no. 3, pp. 621-643, 2003.

[14] R. Jeurissen, and J. van den Berg, "Index tracking using a hybrid genetic algorithm," *Proceedings of 2005 ICSC Congress on Computational Intelligence Methods and Applications*, 2005.

[15] J. Park, D. Yang, and K. Park, "Approximate dynamic programming-based dynamic portfolio optimization for constrained index tracking," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 13, no. 1, pp. 19-28, 2013.

[16] S. Boyd, M. Mueller, B. O'Donoghue, and Y. Wang, "Performance bounds and suboptimal policies for multi-period investment," *Foundations and Trends in Optimization*, vol. 1 no. 1, pp. 1-69, 2014.

[17] J. Primbs, "Portfolio optimization applications of stochastic receding horizon control," *Proceeding of the 2007 American Control Conference*, pp. 1811-1816, 2007.

[18] J. Park, J. Jeong, and K. Park, "An investigation on dynamic portfolio selection problems utilizing stochastic receding horizon approach," *Journal of Korean Institute of Intelligent Systems*, vol. 22, no. 3, pp. 386-393, 2012.

[19] M. Dai, Q. Zhang, and Q. J. Zhu, "Trend following trading under a regime switching model," *SIAM Journal on Financial Mathematics*, vol. 1, pp. 780-810, 2010.

[20] J. Park, D. Yang, and K. Park, "Investigations on dynamic trading strategy utilizing stochastic optimal control and machine learning," *Journal of Korean Institute of Intelligent Systems*, vol. 23, no. 4, pp. 348-353, 2013.

[21] M. Dai, Q. Zhang, and Q. J. Zhu, "Optimal trend following trading rules," (July 19, 2011) Available at SSRN: http://ssrn.com/abstract=1762118 or http://dx.doi.org/10.2139/ssrn.1762118.

[22] H. T. Kong, Q. Zhang, and G. G. Yin, "A trend-following strategy: Conditions for optimality," *Automatica*, vol. 47, no. 4, pp. 661-667, 2011.

[23] J. Yu, and Q. Zhang, "Optimal trend-following trading rules under a three-state regime switching model," *Mathematical Control and Related Fields*, vol. 2, no. 1, pp. 81-100, 2012.

[24] S. Mudchanatongsuk, J. A. Primbs, and W. Wong, "Optimal pairs trading: A stochastic control approach," In *Proceedings of 2008 American Control Conference*, 2008.

[25] J. Hensman, A. G. de G. Matthews, and Z. Ghahramani, "Scalable variational Gaussian process classification," arXiv preprint arXiv:1411.2005, 2014.

[26] Y. Gal, Y. Chen, and Z. Ghahramani, "Latent Gaussian processes for distribution estimation of multivariate categorical data," arXiv preprint arXiv:1503.02182, 2015.

[27] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," arXiv preprint arXiv:1312.6114, 2013.

[28] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic backpropagation and approximate inference in deep generative models," arXiv preprint arXiv:1401.4082, 2014.

[29] J. Park, J. Lim, W. Lee, S. Ji, K. Sung, and K. Park, "Modern probabilistic machine learning and control methods for portfolio optimization," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 14, no. 2, pp.73-83, 2014.

[30] H. Peyrl, F. Herzog, and H. P. Geering, "Numerical solution of the Hamilton-Jacobi-Bellman equation for stochastic optmal control problems," In *Proceedings of 2005 WSEAS Int. Conf. on dynamical systems and control*, pp.489-497, 2005.

[31] T. Glasmachers, T. Schaul, Y. Sun, D. Wierstra, and J. Schmidhuber, "Exponential natural evolution strategies," In *Genetic and Evolutionary Computation Conference (GECCO)*, 2010.

[32] Investment Summary Retrieved March 18, 2016, from http://fount.co/

[33] Betterment Investing Made Better Retrieved March 18, 2016, from https://www.betterment.com/

[34] Investment Management, Online Financial Advisor Wealthfront Retrieved March 18, 2016, from https://www.wealthfront.com/

[35] Online Financial Advisor & Investing Advice FutureAdvisor Retrieved March 18, 2016, from https://www.futureadvisor.com/

Jooyoung Park received his BS in Electrical Engineering from Seoul National University in 1983 and his PhD in Electrical and Computer Engineering from the University of Texas at Austin in 1992. He joined Korea University in 1993, where he is currently a professor at the Department of Control and Instrumentation Engineering. His recent research interests are in the areas of machine learning, control theory, and financial engineering.
E-mail: parkj@korea.ac.kr

Seongman Heo received his BS in Control and Instrumentation Engineering from Korea University in 2015. Currently, he is a graduate student (master course) at Korea University majoring in Control and Instrumentation Engineering. His research areas include control theory, machine learning and deep learning.
E-mail: hsm0099@korea.ac.kr

Taehwan Kim received his BS in Control and Instrumentation Engineering from Korea University in 2015. Currently, he is a graduate student (master course) at Korea University majoring in Control and Instrumentation Engineering. His research areas include control theory, deep learning and reinforcement learning.
E-mail: kteaw0110@korea.ac.kr

Jeongho Park received his BS in Control and Instrumentation Engineering from Korea University in 2016. Currently, he is a graduate student (master course) at Korea University majoring in Control and Instrumentation Engineering. His research areas include control theory, machine learning and pattern recognition.
E-mail: seanpark0107@korea.ac.kr

Jaein Kim received his Bs in Mathematics & Information from Gachon University in 2012. Currently, he is a graduate student at Korea University majoring in Mathemaitics. His research areas include applied mathematics, machine learning and pattern recognition.
E-mail: kkjin85@korea.ac.kr

Kyungwook Park received his BBA and MBA from Seoul National University and his PhD in Finance from the University of Texas at Austin in 1993. He joined Korea University in 1994, where he is currently a professor at the School of Business Administration. His recent research interests are in the areas of derivatives and hedging, control-theory-based assets and derivatives management, and cost of capital estimation with derivative pricing.
E-mail: pkw@korea.ac.kr