6-2017

Markowitz Portfolio Theory Helps Decrease Medicines' Side Effect and Speed Up Machine Learning

Thongchai Dumrongpokaphan
Chiang Mai University, tcd43@hotmail.com

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Comments:
Technical Report: UTEP-CS-17-68

Markowitz Portfolio Theory Helps Decrease Medicines' Side Effect and Speed Up Machine Learning

Thongchai Dumrongpokaphan and Vladik Kreinovich

Abstract In this paper, we show that, similarly to the fact that distributing the investment between several independent financial instruments decreases the investment risk, using a combination of several medicines can decrease the medicines' side effects. Moreover, the formulas for optimal combinations of medicine are the same as the formulas for the optimal portfolio, formulas first derived by the Nobel-prize winning economist H. M. Markowitz. A similar application to machine learning explains a recent success of a modified neural network in which the input neurons are also directly connected to the output ones.

1 Markowitz Portfolio Theory: A Brief Reminder

The main idea behind Markowitz portfolio theory. In his Nobel-prize winning paper [5], H. M. Markowitz proposed a method for selecting an optimal portfolio of financial investments.

To be explain the main ideas behind his method, let us start with a simple case when we have n independent financial instrument, each with a known expected return-on-investment μ_i and a known standard deviation σ_i. In principle, we can combine these portfolios, by allocating the part w_i of our investment amount to the i-th instrument. Here, we have $w_i \geq 0$ and $\sum\limits_{i=1}^{n} w_i = 1$.

For each of these portfolios, we can determine the expected return on investment μ and the standard deviation σ from the formulas

Thongchai Dumrongpokaphan
Faculty of Science, Chiang Mai University, Chiang Mai, Thailand, e-mail: tcd43@hotmail.com

Vladik Kreinovich
Department of Computer Science, University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA, e-mail: vladik@utep.edu

$$\mu = \sum_{i=1}^{n} w_i \cdot \mu_i \text{ and } \sigma^2 = \sum_{i=1}^{n} w_i^2 \cdot \sigma_i^2.$$

Some of such portfolios are less risky – i.e., have smaller standard deviation – but have a smaller expected return on investment. Other portfolios have a larger expected return on investment but are more risky.

We can therefore formulate two possible problems:

- The first problem is when we want to achieve a certain expected return on investment μ. Out of all possible portfolios that provide such expected return on investment, we want to find the portfolio for which the risk σ is the smallest possible.
- The second problem is when we know the maximum amount of risk σ that we can tolerate. There are several different portfolios that provide the allowed of risk. Out of all such portfolios, we would like to select the one that provides the largest possible return on investment.

Example. Let us consider the simplest case, when all n instruments have the same expected return on investment $\mu_1 = \ldots = \mu_n$ and the same standard deviation $\sigma_1 = \ldots = \sigma_n$. In this case, the problem is completely symmetric with respect to permutations, and thus, the optimal portfolio should be symmetric too. Therefore, all the parts must be the same: $w_1 = \ldots = w_n$. Since $\sum\limits_{i=1}^{n} w_i = 1$, this implies that $w_1 = \ldots = w_n = \dfrac{1}{n}$. For these values w_i, the expected return on investment is equal to the same value as for each instrument $\mu = \mu_1$, but the risk decreases:

$$\sigma^2 = \sum_{i=1}^{n} w_i^2 \cdot \sigma_i^2 = n \cdot \frac{1}{n^2} \cdot \sigma_1^2 = \frac{1}{n} \cdot \sigma_1^2,$$

hence $\sigma = \dfrac{\sigma_1}{\sqrt{n}}$.

What we can conclude from this example. A natural conclusion is that if we diversify our portfolio, i.e., if we divide our investment amount between different independent financial instruments, then we can drastically decrease the corresponding risk.

A similar idea works well in measurement. If we have n results $x_1, \ldots, x_n$ of measuring the same quantity x, with measurement error $x_i - x$ with mean 0 and standard deviation σ_i, and if the measurement errors corresponding to different measurements are independent, then we can decrease the estimation error if,

- instead of the original estimates x_i for the quantity x,
- we use their weighted average $\widetilde{x} = \sum\limits_{i=1}^{n} w_i \cdot x_i$, for some weights $w_i \geq 0$ for which

$$\sum_{i=1}^{n} w_i = 1;$$

see, e.g., [6].

In this case, the standard deviation of the estimate $\widetilde{x}$ is equal to

$$\sigma^2 = \sum_{i=1}^{n} w_i^2 \cdot \sigma_i^2.$$

We want to find the weights w_i that minimize σ^2 under the given constraint $\sum\limits_{i=1}^{n} w_i = 1$. By using the Lagrange multiplier method, we can reduce this constraint optimization problem to the following unconstrained optimization problem:

$$\sum_{i=1}^{n} w_i^2 \cdot \sigma_i^2 + \lambda \cdot \left(\sum_{i=1}^{n} w_i - 1 \right) \to \min_i.$$

Differentiating the resulting objective function with respect to w_i and equating the derivative to 0, we conclude that $2w_i \cdot \sigma_i^2 + \lambda = 0$, thus, $w_i = c \cdot \sigma_i^{-1}$, for some constant $c \stackrel{\text{def}}{=} -\dfrac{\lambda}{2}$. This constant c can be found from the condition that $\sum\limits_{i=1}^{n} w_i = 1$: we get $c = \dfrac{1}{\sum\limits_{j=1}^{n} \sigma_j^{-2}}$ and thus,

$$w_i = \frac{\sigma_i^{-2}}{\sum\limits_{j=1}^{n} \sigma_j^{-2}}.$$

For these weights, we get

$$\sigma^2 = \sum_{i=1}^{n} w_i^2 \cdot \sigma_i^2 = \sum_{i=1}^{n} \frac{\sigma_i^{-4}}{\left(\sum\limits_{j=1}^{n} \sigma_j^{-2} \right)^2} \cdot \sigma_i^2 = \sum_{i=1}^{n} \frac{\sigma_i^{-2}}{\left(\sum\limits_{j=1}^{n} \sigma_j^{-2} \right)^2} =$$

$$\frac{\sum\limits_{i=1}^{n} \sigma_j^{-2}}{\left(\sum\limits_{j=1}^{n} \sigma_j^{-2} \right)^2} = \frac{1}{\sum\limits_{j=1}^{n} \sigma_j^{-2}}.$$

The sum $\sum\limits_{j=1}^{n} \sigma_j^{-2}$ is larger than each of its terms σ_j^{-2}, and thus, the inverse σ^2 of this sum is smaller than each of the inverses σ_j^2. So, combining measurement results indeed decreases the approximation error.

In particular, when all measurements are equally accurate, i.e., when $\sigma_1 = \ldots = \sigma_n$, we get $\sigma = \dfrac{\sigma}{\sqrt{n}}$.

Optimal portfolio when different instruments are independent. In the previous text, we considered the case when different financial instruments are independent and identical. Let us now consider a more general case, when we still assume that the financial instruments are independent, but we take into account that these instrument are, in general, different, i.e., they have individual values μ_i and σ_i.

In this case, the first portfolio optimization problem takes the following form: minimize

$$\sum_{i=1}^n w_i^2 \cdot \sigma_i^2$$

under the constraints

$$\sum_{i=1}^n w_i \cdot \mu_i = \mu \text{ and } \sum_{i=1}^n w_i = 1.$$

For this problem, Lagrange multiplier methods leads to minimizing the expression

$$\sum_{i=1}^n w_i^2 \cdot \sigma_i^2 + \lambda \cdot \left(\sum_{i=1}^n w_i \cdot \mu_i - \mu\right) + \lambda' \cdot \left(\sum_{i=1}^n w_i - 1\right).$$

Differentiating this expression with respect to w_i and equating the derivative to 0, we conclude that

$$2w_i \cdot \sigma_i^2 + \lambda \cdot \mu_i + \lambda' = 0,$$

i.e., that

$$w_i = a \cdot (\mu_i \cdot \sigma_i^{-2}) + b \cdot \sigma_i^{-2},$$

where $a \stackrel{\text{def}}{=} -\dfrac{\lambda}{2}$ and $b \stackrel{\text{def}}{=} -\dfrac{\lambda'}{2}$. For these values w_i, the constraints $\sum\limits_{i=1}^n w_i \cdot \mu_i = \mu$ and $\sum\limits_{i=1}^n w_i = 1$ take the form

$$a \cdot \Sigma_2 + b \cdot \Sigma_1 = \mu; \text{ and } a \cdot \Sigma_1 + b \cdot \Sigma_0 = 1,$$

where we denoted $\Sigma_k \stackrel{\text{def}}{=} \sum\limits_{i=1}^n (\mu_i)^k \cdot \sigma_i^{-2}$. Thus,

$$a = \frac{\Sigma_1 - \mu \cdot \Sigma_0}{\Sigma_1^2 - \Sigma_0 \cdot \Sigma_2} \text{ and } b = \frac{\mu \cdot \Sigma_1 - \Sigma_2}{\Sigma_1^2 - \Sigma_0 \cdot \Sigma_2}.$$

General case. In general, we may have correlations ρ_{ij} between different financial instruments. In this case, the standard deviation of the weighted combination has the form

$$\sum_{i=1}^n w_i^2 \cdot \sigma_i^2 + \sum_{i \neq j} \rho_{ij} \cdot w_i \cdot w_j \cdot \sigma_i \cdot \sigma_j.$$

This is a quadratic function, thus the Lagrange multiplier form is also quadratic, and after differentiating it and equating the derivatives to 0 we get an easy-to-solve system of linear equations.

2 How Markowitz Portfolio Theory Can Be Applied to Medicine

Formulation of the problem in informal terms. In medicine, usually, for each disease, we have several possible medicines. All these medicines are usually reasonable effective – otherwise they would not have been approved by the corresponding regulatory agency – but all of them usually have some undesirable side effects. How can we decrease these side effects?

A natural idea. The example of portfolio optimization prompts a natural idea: instead of applying individual medicines, try a combination of several medicines.

To see whether this approach will indeed work, let us reformulate our problem in precise terms.

Let us reformulate this problem in precise terms. We want to change the state of the patient: to bring the patient from a sick state to the healthy state. Each state can be described by the values of all the parameters that characterize this state: body temperature, blood pressure, etc.

We want to move the patient from the current sick state $s = (s_1, \ldots, s_d)$ to the desired healthy state $h = (h_1, \ldots, h_d)$.

We want to describe the joint effect of taking several medicines. Let us measure the dose w_i of each medicine i by considering the proportion to the actual dose to usually prescribed dose. In these units, the usually prescribed dose is $w_i = 1$. Let us describe the state of a patient after taking the doses $w = (w_1, \ldots, w_n)$ of different medicines by $f(w)$.

When no medicines are applied, i.e., when $w_i = 0$ for all i, then the patient remains sick, in the state s: $f(0) = s$. Doses of medicine are usually reasonable small, to avoid harmful side effects – we are not talking about life-and-death situations where strong measures are applied and side effects (like crushed ribs during the heart massage) are a price everyone is willing to pay to stay alive. Since the doses are small, we can expand the dependence $f(w)$ of the state on the doses w_i in Taylor series and keep only linear terms in this dependence; taking into account that $f(0) = s$, we conclude that

$$f(w) = s + \sum_{i=1}^{n} w_i \cdot a_i$$

for some vectors a_i.

We can use this formula to find the resulting state in situation when we apply the full usual dose of the i-th medicine, i.e., when we take $w_i = 1$ for this i and $w_{i'} = 0$ for all $i' = i$. In this situation, the resulting state is equal to $s + a_i$. In the ideal world, we should get the state h, i.e., we should have $a_i = h - s$, but in reality, we have side effects, i.e., deviations from this state: $\Delta a_i \stackrel{\text{def}}{=} a_i - (h - s) \neq 0$.

Let σ_i^2 denote the mean square values of this deviation Δa_i. Substituting the expression $a_i = (h-s) + \Delta a_i$ into the formula for the resulting state $f(w)$, we conclude that the joint effect of several medicine is equal to

$$f(w) = s + \sum_{i=1}^{n} w_i \cdot (h-s) + \sum_{i=1}^{n} w_i \cdot \Delta a_i.$$

We want to make sure that, modulo side effects, we get into the healthy state h, i.e., that $s + (h-s) \cdot \sum\limits_{i=1}^{n} w_i = h$. This condition is equivalent to $\sum\limits_{i=1}^{n} w_i = 1$.

Under this condition, we want to minimize the overall side effect, i.e., we want to minimize its mean squared value. When all medicines are different, side effects are independent, and thus, for the mean square error σ of the overall side effect, we have the formula $\sum\limits_{i=1}^{n} w_i^2 \cdot \sigma_i^2$.

Thus, to get the optimal combination of medicines, we must find, among all the values w_i for which $\sum\limits_{i=1}^{n} w_i = 1$, the combination that minimizes the sum $\sum\limits_{i=1}^{n} w_i^2 \cdot \sigma_i^2$.

This is exactly Markowitz formula. The above optimization problem is exactly the Markowitz problem – with $\mu_i = 1$. This is also the exact same problem as we encountered when combining different independent measurement results. Thus, we conclude that we should take $w_i = \dfrac{\sigma_i^{-2}}{\sum\limits_{j=1}^{n} \sigma_j^{-2}}$. This will enable us to decrease the side effects to the level $\sigma^2 = \dfrac{1}{\sum\limits_{j=1}^{n} \sigma_j^{-2}}$.

In particular, in situations when all the medicines are of approximate the same quality, i.e., when all side effects are of the same strength $\sigma_1 = \ldots = \sigma_n$, we should take all the medicines with equal weight $w_1 = \ldots = w_n = \dfrac{1}{n}$. This will enable us to decrease the side effects to the level $\sigma = \dfrac{\sigma_1}{n}$.

What if side effects are correlated. The above analysis assumes that all side effects are independent. In reality, side effects may be correlated. It is therefore desirable to take this correlation into account.

In the symmetric case, when $\sigma_1 = \ldots = \sigma_n$, even if we allow the possibility of correlations - but assume that correlation is approximately the same for all pairs of medicines $\rho_{ij} \approx \rho$ – due to symmetry, we will still get the optimal combination in which each medicine is taken in the same dose $w_1 = \ldots = w_n = \dfrac{1}{n}$. The only difference is that if there is a correlation, the decrease in side effects will be not as drastic as in the independent case. Namely, we will have

$$\sigma^2 = \sum_{i=1}^{n} w_i^2 \cdot \sigma_i^2 + \sum_{i \neq j} \rho_{ij} \cdot w_i \cdot w_j \sigma_i \cdot \sigma_j =$$

$$n \cdot \frac{1}{n^2} \cdot \sigma_1^2 + n \cdot (n-1) \cdot \rho \cdot \frac{1}{n^2} \cdot \sigma_1^2 = \sigma_1^2 \cdot \left(\rho + \frac{1-\rho}{n} \right).$$

This decrease in side effects has actually been experimentally observed. Recent analysis of experimental data shows that for hypertension, a combination of quarter-doses of four different medicines indeed drastically decreases the corresponding side effect [1, 4] – so this is real!

3 Applications to Machine Learning

Description of the problem. In many cases, when the inputs are small, we can use linear models – just as we did in medical applications. When the inputs are large, linear models often no longer work, and we often do not know what type of non-linear dependence we have. To describe such dependencies, we can use machine learning techniques that allow us to approximate any possible non-linear dependencies; see, e.g., [2].

In the intermediate case, we can use both models:

- we can use a linear model, and
- we can also use machine learning techniques – such as neural networks.

Both models are not perfect: linear models are not very accurate while machine learning models are much more accurate but require a lot of time to train. Can we combine the advantages of these models?

Markowitz-motivated idea. Instead of considering the estimate $f_{\text{NN}}(x)$ generated by a neural network and a linear model $f_{\text{lin}}(x) = a_0 + \sum\limits_{i=1}^{n} a_i \cdot x_i$, let us consider the weighted combinations of these models, i.e., functions of the type

$$f(x) = w_{\text{NN}} \cdot f_{\text{NN}}(x) + b_0 + \sum_{i=1}^{n} b_i \cdot x_i,$$

where we denoted $b_i = w_{\text{lin}} \cdot a_i = (1 - w_{\text{NN}}) \cdot a_i$.

This idea also works! It turns out that this idea can indeed drastically speed up the neural networks, see [3].

Interestingly, the addition of linear terms did not even require big changes in the training algorithm. Indeed, usually, neural networks have:

- an intermediate layer, where the input signals $x_1, \ldots, x_n$ undergo some nonlinear transformations into values $z_k = f_k(x_1, \ldots, x_k)$, followed by

- the output layer, where linear neurons transforms the values z_k coming from the intermediate layer into the final outputs $y = f_{\rm NN}(x) = \sum\limits_k W_k \cdot z_k - W_0$.

To incorporate additional linear terms $b_i \cdot x_i$, all we need to do is to add direct connections from the input layer to the output layer. This way, the signal produced by the output neuron is a linear combination of the signals z_k from the intermediate layer *and* the inputs x_i and thus, has the form

$$y = \sum_k W_k \cdot z_k - W_0 + \sum_{i=1}^n b_i \cdot x_i,$$

i.e., the desired form $y = f_{\rm NN}(x) + \sum\limits_{i=1}^n b_i \cdot x_i$, where we denoted

$$f_{\rm NN}(x) \stackrel{\rm def}{=} \sum_k W_k \cdot z_k - W_0 = \sum_k W_k \cdot f_k(x_1, \ldots, x_n) - W_0.$$

This minor change in the structure of a neural network still allows us to use practically the same standard computationally backpropagation algorithm (see, e.g., [2]) for training – after a very small and computationally insignificant modification.

Acknowledgments

This work was supported by Chiang Mai University, Thailand. It was also supported in part by NSF grant HRD-1242122.

One of the authors (VK) is greatly thankful to Philip Chen for valuable discussions.

References

1. A. Bennett, C. K. Chow, M. Chou, H.-M. Dehbi, R. Webster, A. Salam, A. Patel, B. Neal, D. Peiris, J. Thakkar, J. Chalmers, M. Nelson, C. Reid, G. S. Hillis, M. Woodward, S. Hilmer, T. Usherwood, S. Thom, and A. Rodgers, "Efficacy and Safety of Quarter-Dose Blood PressureLowering Agents: A Systematic Review and Meta-Analysis of Randomized Controlled Trials", *Hypertension*, 2017, Vol. 69, No. 6, pp. 85–93.
2. C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, 2006.
3. S. Feng and C. L. P. Chen, "A fuzzy restricted boltzmann machine: novel learning algorithms based on crisp possibilistic mean value of fuzzy numbers", *IEEE Transactions on Fuzzy Systems*, 2017, to appear.
4. G. Grassi and G. Mancia, "Quarter dose combination therapy: good news for blood pressure control", *Hypertension*, 2017, Vol. 69, No. 6, pp. 32–34.
5. H. M. Markowitz, "Portfolio selection", *The Journal of Finance*, 1952, Vol. 7, No. 1, pp. 77–91.
6. S. G. Rabinovich, *Measurement Errors and Uncertainty: Theory and Practice*, Springer Verlag, Berlin, 2005.