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

Thongchai Dumrongpokaphan[1] and Vladik Kreinovich[2]

[1]Faculty of Science, Chiang Mai University
Chiang Mai, Thailand, tcd43@hotmail.com
[2]University of Texas at El Paso, USA, vladik@utep.edu

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

1. The Main Idea Behind Markowitz Portfolio Theory

- In his Nobel-prize winning paper, H. M. Markowitz proposed a method for selecting an optimal portfolio.
- To 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 with a known standard deviation σ_i.
- We can combine these instruments, by allocating the part w_i of our investment to the i-th instrument.
- Here, we have $w_i \geq 0$ and $\sum\limits_{i=1}^{n} w_i = 1$.

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

2. Markowitz Portfolio Theory (cont-d)

- For each portfolio, we can determine the expected return on investment μ and the standard deviation σ:

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

- Some of such portfolios are less risky – i.e., have smaller σ – but have a smaller μ.
- Other portfolios have a larger expected return on investment but are more risky.
- We can therefore formulate two possible problems.

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

3. Markowitz Portfolio Theory (cont-d)

- 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 amount of risk;
  - out of all such portfolios,
  - we would like to select the one that provides the largest possible return on investment.

4. Example

- Let us consider the simplest case, when all n instruments have the same μ and σ:

$$\mu_1 = \ldots = \mu_n, \quad \sigma_1 = \ldots = \sigma_n.$$

- In this case, the problem is completely symmetric with respect to permutations.
- Thus, the optimal portfolio should be symmetric too.
- So, 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 the same $\mu = \mu_i$, but the risk decreases:

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

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

5. 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.

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

6. A Similar Idea Works Well in Measurement

- Suppose that we have n results $x_1, \ldots, x_n$ of measuring the same quantity x.
- Suppose that the measurement error $x_i - x$ has mean 0 and standard deviation σ_i.
- Suppose that measurement errors corresponding to different measurements are independent.
- Then we can decrease the estimation error if,
  - instead of the original estimates x_i for 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 $\sum\limits_{i=1}^{n} w_i = 1$.
- In this case, the standard deviation of the estimate $\widetilde{x}$ is equal to $\sigma^2 = \sum\limits_{i=1}^{n} w_i^2 \cdot \sigma_i^2$.

7. Measurement (cont-d)

- 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 get 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 with respect to w_i and equating the derivative to 0, we get

$$2w_i \cdot \sigma_i^2 + \lambda = 0, \text{ so } w_i = c \cdot \sigma_i^{-1}, \text{ where } c \stackrel{\text{def}}{=} -\frac{\lambda}{2}.$$

- This constant c can be found from the condition that $\sum\limits_{i=1}^{n} w_i = 1$: we get $c = \dfrac{1}{\sum\limits_{j=1}^{n} \sigma_j^{-2}}$; thus, $w_i = \dfrac{\sigma_i^{-2}}{\sum\limits_{j=1}^{n} \sigma_j^{-2}}$.

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

8. Measurement (final)

- For these weights, $\sigma^2 = \sum\limits_{i=1}^n w_i^2 \cdot \sigma_i^2 = \dfrac{1}{\sum\limits_{j=1}^n \sigma_j^{-2}}$.
- The sum $\sum\limits_{j=1}^n \sigma_j^{-2}$ is larger than each of its terms σ_j^{-2}.
- 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 $\sigma = \dfrac{\sigma}{\sqrt{n}}$.

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

9. Optimal Portfolio When Different Instruments Are Independent

- So far, 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 may have individual values μ_i and σ_i.
- In this case, the first portfolio optimization problem takes the following form:
  - minimize $\sum\limits_{i=1}^n w_i^2 \cdot \sigma_i^2$
  - under the constraints $\sum\limits_{i=1}^n w_i \cdot \mu_i = \mu$ and $\sum\limits_{i=1}^n w_i = 1$.

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

10. When Different Are Independent (cont-d)

- Lagrange multiplier methods leads to:

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

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

$$2w_i \cdot \sigma_i^2 + \lambda \cdot \mu_i + \lambda' = 0, \text{ i.e., } w_i = a \cdot (\mu_i \cdot \sigma_i^{-2}) + b \cdot \sigma_i^{-2},$$

$$\text{where } a \stackrel{\text{def}}{=} -\frac{\lambda}{2} \text{ and } b \stackrel{\text{def}}{=} -\frac{\lambda'}{2}.$$

- For these w_i, $w_i \cdot \mu_i = \mu$ and $\sum\limits_{i=1}^{n} w_i = 1$ are:

$$a \cdot \Sigma_2 + b \cdot \Sigma_1 = \mu, \;\; a \cdot \Sigma_1 + b \cdot \Sigma_0 = 1, \text{ where } \Sigma_k \stackrel{\text{def}}{=} \sum_{i=1}^{n} (\mu_i)^k \cdot \sigma_i^{-2}.$$

- Thus, $a = \dfrac{\Sigma_1 - \mu \cdot \Sigma_0}{\Sigma_1^2 - \Sigma_0 \cdot \Sigma_2}$ and $b = \dfrac{\mu \cdot \Sigma_1 - \Sigma_2}{\Sigma_1^2 - \Sigma_0 \cdot \Sigma_2}$.

11. 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.
- After differentiating it and equating the derivatives to 0 we get an easy-to-solve system of linear equations.

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

12. How Markowitz Portfolio Theory Can Be Applied to Medicine

- 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.
- However, all of them usually have some undesirable side effects.
- How can we decrease these side effects?

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

13. 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.

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

14. 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.

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

15. Medical Problem (cont-d)

- Let us measure the dose w_i of each medicine i as a ratio
$$w_i = \frac{\text{actual dose}}{\text{usually prescribed dose}}.$$
- 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: $f(0) = s$.
- Doses of medicine are usually reasonable small, to avoid harmful side effects.

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

16. Medical Problem (cont-d)

- 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 we 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 $f(w) = s + \sum\limits_{i=1}^{n} w_i \cdot a_i$ for some vectors a_i.
- When we apply the full usual dose of the i-th medicine, we get $s + a_i$.

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

17. Medical Problem (cont-d)

- In the ideal world, we should get the state h, i.e., we should have $a_i = h - s$.
- However, 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 Δa_i.
- Substituting $a_i = (h - s) + \Delta a_i$ into the formula for $f(w)$, we get the joint effect of several medicine:

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

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

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

18. Medical Problem (final)

- Under this condition, we want to minimize the overall side effect, i.e., its mean squared deviation.
- When all medicines are different, side effects are independent.
- Thus, for the mean square error σ of the overall side effect, we have the formula $\sum_{i=1}^{n} w_i^2 \cdot \sigma_i^2$.
- Thus, to get the optimal combination of medicines, we must find:
  - among all the values w_i for which $\sum_{i=1}^{n} w_i = 1$,
  - the combination that minimizes the sum $\sum_{i=1}^{n} w_i^2 \cdot \sigma_i^2$.

The Main Idea Behind ...
Example
A Similar Idea Works ...
How Markowitz ...
What If Side Effects ...
Applications to ...
This Idea Also Works!
References to a ...
Reference to a Neural ...

19. This Is Exactly Markowitz Formula

- The above optimization problem is exactly the Markowitz problem – with $\mu_i = 1$.
- We encountered the same problem when combining independent measurement results.
- Thus, we should take $w_i = \dfrac{\sigma_i^{-2}}{\sum\limits_{j=1}^{n} \sigma_j^{-2}}$; this decrease the side effects to the level $\sigma^2 = \dfrac{1}{\sum\limits_{j=1}^{n} \sigma_j^{-2}}$.

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

20. This Is Exactly Markowitz Formula (cont-d)

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

21. 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 still get the optimal combination in which all w_i are equal:

$$w_1 = \ldots = w_n = \frac{1}{n}.$$

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

22. What If Side Effects Are Correlated (cont-d)

- The only difference is that the decrease in side effects may be not as drastic as in the independent case.
- Namely, we will have

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

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

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

23. 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.
- So this is real!

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

24. Applications to Machine Learning

- 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.
- Oftne, we do not know what type of non-linear dependence we have.
- To describe such dependencies, we can use machine learning techniques.
- This allows us to approximate any possible non-linear dependencies.
- 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.

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

25. Applications to Machine Learning (cont-d)

- 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?

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

26. Markowitz-Motivated Idea

- We have an estimate $f_{\rm NN}(x)$ generated by a neural network.
- We also have a linear model $f_{\rm lin}(x) = a_0 + \sum\limits_{i=1}^n a_i \cdot x_i$.
- Let us consider the weighted combinations of these models: $f(x) = w_{\rm NN} \cdot f_{\rm NN}(x) + b_0 + \sum\limits_{i=1}^n b_i \cdot x_i$, where

$$b_i = w_{\rm lin} \cdot a_i = (1 - w_{\rm NN}) \cdot a_i.$$

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

27. This Idea Also Works!

- It turns out that this idea can indeed drastically speed up the neural networks.
- 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:

$$z_k = f_k(x_1, \ldots, x_k),$$

  - followed by the output layer, where linear neurons transforms z_k into the final outputs

$$y = f_{\rm NN}(x) = \sum_k W_k \cdot z_k - W_0.$$

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

28. This Idea Also Works (cont-d)

- 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.
- Then, $y = \sum_k W_k \cdot z_k - W_0 + \sum_{i=1}^{n} b_i \cdot x_i$.
- Thus, $y = f_{\rm NN}(x) + \sum_{i=1}^{n} b_i \cdot x_i$, where

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

- This minor change in the neural network still allows us to use the standard backpropagation algorithm.

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

29. 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.

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

30. References to a Medical Success

- 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 Pressure-Lowering Agents: A Systematic Review and Meta-Analysis of Randomized Controlled Trials", *Hypertension*, 2017, Vol. 69, No. 6, pp. 85–93.
- G. Grassi and G. Mancia, "Quarter dose combination therapy: good news for blood pressure control", *Hypertension*, 2017, Vol. 69, No. 6, pp. 32–34.

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .

31. Reference to a Neural Network Success

- 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.

The Main Idea Behind . . .
Example
A Similar Idea Works . . .
How Markowitz . . .
What If Side Effects . . .
Applications to . . .
This Idea Also Works!
References to a . . .
Reference to a Neural . . .