機械学習とポートフォリオ選択の素敵な関係

武田　朗子，後藤　順哉

本稿では「機械学習」と「金融工学」という，独立に生まれ，発展してきた 2 つの分野の方法論に共通する構造と差異に着目して，新しい発見に繋げようという試みを紹介したい．とりわけ，「機械学習」と「金融工学」の間では双方向への貢献が可能であること，そして両分野間に素敵な関係を見いだすことができることを示す．

キーワード：機械学習，線形回帰，サポートベクターマシン，ポートフォリオ選択，金融リスク指標

1. はじめに

数理モデルを扱っていると，異なる対象でありながら，同じような数理的構造を有した問題に遭遇することが少なくない．実際，「この問題は○○問題に帰着された」といった表現はその最たる例である．一方で，「同じよう」であっても「全く同じ」ことは稀である．その場合，その差分を検討してみることで素敵な発見が生まれることもある．

本稿では「機械学習」と「金融工学」という，独立に生まれ，発展してきた 2 つの分野の方法論に共通する構造と差異に着目して，新しい発見に繋げようという試みを，筆者らの研究成果を中心に 2 つ紹介したい（図 1）．

- 金融リスク尺度を用いてサポートベクターマシンのモデルの妥当性を理論的に示した，「金融工学手法の機械学習への適用」の研究．
- 機械学習の “正則化” をポートフォリオ最適化モデルに取り入れた，「機械学習手法の金融工学への適用」の研究．

これらの研究成果は「機械学習」と「金融工学」の間では双方向への貢献が可能であることを示唆しており，両分野間に素敵な関係を見いだすことができる．

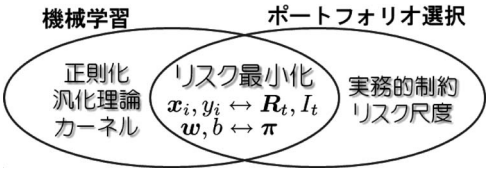

図 1　機械学習とポートフォリオ選択の関係

たけだ　あきこ
慶應義塾大学理工学部管理学科
〒 223–8522 横浜市港北区日吉 3–14–1
ごとう　じゅんや
中央大学理工学部経営システム工学科
〒 112–8551 文京区春日 1–13–27

まず，機械学習の方法論の概略から始めよう．

2. 機械学習の基本戦略

機械学習（machine learning）を簡単に言うと，「見えている情報（データ）を手がかりに，見えていないものを予測・発見する仕組みを獲得するための技術」である．数値・文字・画像・音声など多種多様なデータの中から，規則性・パターン・知識を発見し，現状の把握や将来の予測をするのにその知識を役立てるのが機械学習の目的である．

本稿では，**サポートベクターマシン**（**SVM**）を中心に，いくつかの機械学習手法を取りあげる．SVM は判別分析や回帰分析，外れ値検出など，さまざまな方法の総称である．本稿では，変数が n 個の線形回帰問題を中心に説明したい．入力と出力の組：$(\boldsymbol{x}_i, y_i)$, $i \in M := \{1, \ldots, m\}$ が与えられており，入力 $\boldsymbol{x}_i$ は n 次元の実数ベクトル（つまり，$\boldsymbol{x}_i \in \mathbb{R}^n$），出力 y_i は実数値とする．“学習” とは，これらのデータに “何らかの基準” で最も合う関数関係 $y = h(\boldsymbol{x})$ を求めることである（図 2 を参照のこと）．このような関数関係 $y = h(\boldsymbol{x})$ が求められれば，未知のデータ $\boldsymbol{x}$ に関数を適用することで，予測 $y = h(\boldsymbol{x})$ を与えることができる[1]．

[1] 例えば，立地（駅までの距離や商店街までの距離），築年数，部屋数，敷地の広さ，その他の要因に基づいて住宅価格を予測する例を考えてみよう．駅までの距離や築年数などを入力 $\boldsymbol{x}$ としてベクトルで表現し，出力を住宅価格 y とする．一定期間に観測した結果，住宅価格 y_i と住宅のデータ $\boldsymbol{x}_i$ が多数（m 個）集まったとする．これらのデータ：$(\boldsymbol{x}_i, y_i)$, $i \in M$，に基づいて予測式 $h(\boldsymbol{x})$ を作成することにより，今後，新しいデータ $\boldsymbol{x}$ に対して，住宅価格 $y = h(\boldsymbol{x})$ の予測ができる．

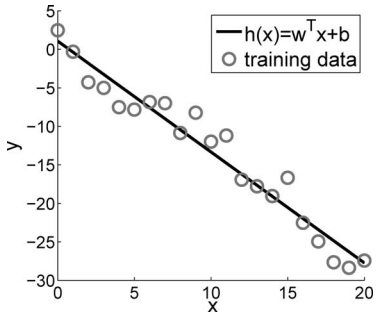

図 2　直線へのあてはめ

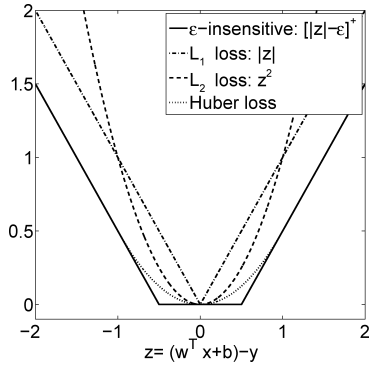

図 3　損失関数（Huber 損失のパラメータは $\mu = 1$, ϵ 許容損失のパラメータは $\epsilon = 0.5$）

2.1　損失関数と正則化

どのような基準を用いて，最も合う関数関係 $y = h(\boldsymbol{x})$ を求めればよいだろうか．推定値と実測値の差，つまり残差は $h(\boldsymbol{x}) - y$ と表わされる．線形回帰モデルであれば，残差は $z(\boldsymbol{x}, y) := h(\boldsymbol{x}) - y = (\boldsymbol{w}^\top \boldsymbol{x} + b) - y$ となる．ここで，推定の悪さを定義した関数（損失関数と呼ぶ）を $L(h, z(\boldsymbol{x}, y))$，もしくは誤解を与えない範囲内で省略して $L(h, z)$ と表わすことにする．

回帰モデルの目的は，未知の確率分布 $P(\boldsymbol{x}, y)$ に独立に従う訓練データ（モデルを作るためのデータ）を用いて，**汎化誤差**（学習に使わなかった未知のデータに対する予測誤差）：

$$R[h] := \int L(h, z(\boldsymbol{x}, y))\, dP(\boldsymbol{x}, y)$$

を最小化する h を推定することである[2]．

ここで L としては以下のようなさまざまな候補が考えられる．

- ϵ 許容損失 (ϵ-insensitive loss): $L(h, z) = [|z| - \epsilon]^+ := \max\{0, |z| - \epsilon\}$．パラメータ ϵ に対して，$|z| \le \epsilon$ の損失は無視するが，その外側の範囲では線形に増加する関数．サポートベクター回帰（後で記す ν-SVR など）で用いられる損失関数．
- 絶対損失 (ℓ_1 loss): $L(h, z) = |z|$.
- 二乗損失 (ℓ_2 loss)：$L(h, z) = z^2$.
- Huber 損失：パラメータ μ に対して，$|z| \le \mu$ の範囲では 2 次関数だが，その外側の範囲では線形に増加する関数．

図 3 はこれらの損失関数 L を図示したものである．損失関数にはそれぞれ特徴がある．例えば，Huber 損失をはじめとする，z が 0 から遠いところでは線形にしか増加しない関数は，外れ値の影響が 2 次の損失関数と比べてずっと小さいため，データの外れ値の影響を受けにくい，といった利点がある．異なる損失関数を選択すれば異なる回帰モデルが得られる．しかしながら，いずれの損失関数においても「小さな損失は受け入れるが大きな損失は避けたい」という考えを念頭において設計されている．

確率分布 $P(\boldsymbol{x}, y)$ は未知であり，実際には，$R[h]$ を最小化する h を求めることはできない．また，データ $D := \{(\boldsymbol{x}_i, y_i) : i \in M\}$ が与えられたときの経験損失：

$$L_{emp}(h, D) = \frac{1}{m} \sum_{i=1}^{m} L(h, (\boldsymbol{x}_i, y_i))$$

だけを最適化する関数 h を求めても，過適合（訓練データに対して学習されているが未知データに対しては有用な情報が得られていないこと）のため汎化誤差は最小にならない．

こうした場合には，平滑化などを行う罰則項（以降，正則化項と呼ぶ）$\Omega(h)$ と正値のパラメータ λ もしくは C を用いて次のような最小化問題を考える．

$$\min_h L_{emp}(h, D) + \lambda\Omega(h), \text{ もしくは}$$
$$\min_h L_{emp}(h, D) \text{ s.t. } \Omega(h) \le C$$

こうした問題の書き換えを**正則化** (regularization) という．これら 2 つの問題は正則化項を目的関数に加えるか，制約式として表わすかの違いがあるものの，λ と C を対応させて設定すれば，双方とも同じ最適解をもつ等価な問題である．

正則化項 $\Omega(h)$ としては，ℓ_2 ノルムを用いた $\|\boldsymbol{w}\|_2$ や ℓ_1 ノルムを用いた $\|\boldsymbol{w}\|_1$ が代表的である．二乗損失に ℓ_2 ノルムの正則化項を加えた式を最小化するモデルを ridge（リッジ）回帰といい，ℓ_1 ノルムの正則化

[2] 先程の住宅価格の例を挙げて説明すると，過去のデータについてではなく，購入を検討している住宅の “本当の価格” をどれだけ言い当てられるかを定量化したものが汎化誤差である．

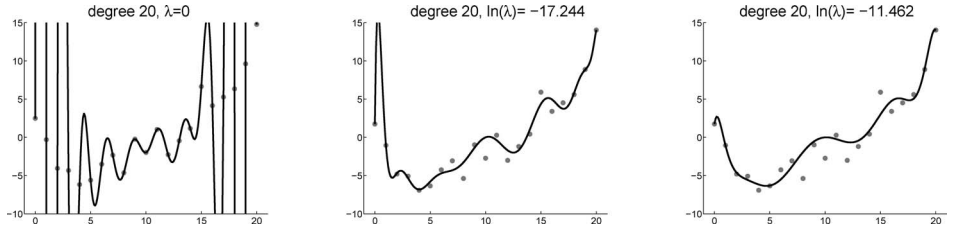

図 4 ridge 回帰において，ℓ_2 正則化項のパラメータ λ が $d(=20)$ 次の多項式：$h(x) = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + \cdots + w_d x^d$ へのあてはめに及ぼす影響

項を加えた回帰モデルを lasso（ラッソ）回帰と呼ぶ．ridge 回帰は簡単に解析解が求まるという利点がある．一方，lasso 回帰を用いると，0 になる w_i が多い，疎な解が得られやすいと言われ，特徴選択（n 個の属性から成る入力 $\boldsymbol{x} \in \mathbb{R}^n$ に対して，n 個の属性をすべて使わず，その中で意味のある部分集合だけを選択する手法）に役立つことが知られている．

最近は，より直接的に疎な解を得るための正則化項として，$\|\boldsymbol{w}\|_0$ や $\|\boldsymbol{w}\|_{1/2}$ を用いた回帰モデルも目にする[3]．また $\|\boldsymbol{w}\|_1$ と $\|\boldsymbol{w}\|_2$ を正値パラメータ 2 つ (λ_1, λ_2) を用いて組み合わせた elastic net [13] と呼ばれる正則化項 $\lambda_1\|\boldsymbol{w}\|_1 + \lambda_2\|\boldsymbol{w}\|_2$ も提案されている．

図 4（左）が示すように，二乗損失最小化により 20 次の多項式に訓練データをあてはめようとすると，過適合を起こしてしまう．しかし，ℓ_2 正則化項を加えて，パラメータ λ を 0 から大きくすることにより，過適合を防ぎ，滑らかな曲線を得ることができる．左図の曲線よりも右図の曲線のほうが，新しいデータに対して予測能力は高い．

過適合を防ぐために正則化の工夫を行い，汎化誤差をいかに小さくするかが，ここで扱う機械学習の基本戦略である．

2.2 金融工学→機械学習

金融工学のトピックの 1 つとして，分散に代わるリスク尺度の概念の再検討がなされてきた．代表的代替案の 1 つが value-at-risk (**VaR**) であり，条件付き VaR (conditional VaR, **CVaR**) である．確率分布の言葉を借りれば，前者は損失分布のパーセント点，後者はそれを上回る損失の（条件付き）期待値である．

[3] ここで，$\|\cdot\|_\delta$ $(\delta > 0)$ は ℓ_δ ノルムであり，$\boldsymbol{w} \in \mathbb{R}^n$ に対して，

$$\|\boldsymbol{w}\|_\delta := \left[\sum_{i=1}^{n} |w_i|^\delta\right]^{(1/\delta)}$$

で定義される．この定義による ℓ_δ ノルムは $\delta < 1$ に対しては数学的な意味での “ノルム” ではないことに注意．また，形式的に ℓ_0 ノルムを $\|\boldsymbol{x}\|_0 := \lim_{\delta \to 0+} \|\boldsymbol{x}\|_\delta$ =「ベクトル $\boldsymbol{w}$ 中の非 0 成分の数」と定義する．

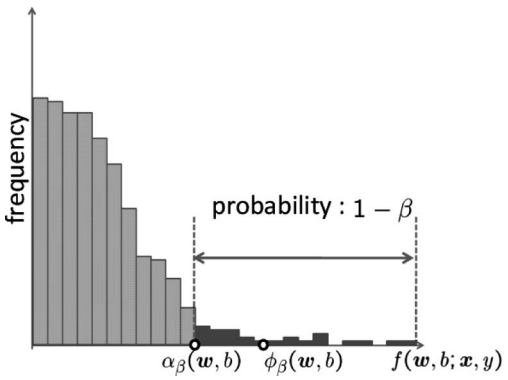

図 5 $f(\boldsymbol{w}, b; \boldsymbol{x}, y) := |(\boldsymbol{w}^\top \boldsymbol{x}_i + b) - y_i|$, $i \in M$ のヒストグラム

そのアイデアを回帰分析に適用してみよう．絶対損失の経験分布 $f(\boldsymbol{w}, b; \boldsymbol{x}_i, y_i) := |(\boldsymbol{w}^\top \boldsymbol{x}_i + b) - y_i|$, $i \in M$, を考える．この絶対損失は小さければ小さいほど良い．

$\beta \in [0, 1)$ に対して，$\alpha_\beta(\boldsymbol{w}, b)$ をこの損失分布の 100β-パーセント点として定義する．つまり，$f(\boldsymbol{w}, b; \boldsymbol{x}_i, y_i)$ のうち，$100(1-\beta)\%$のサンプルだけが閾値 $\alpha_\beta(\boldsymbol{w}, b)$ よりも大きな値をとることになる（図 5）．$\alpha_\beta(\boldsymbol{w}, b)$ は金融分野では VaR として知られ，ポートフォリオのリスクを計る指標としてしばしば使われている．

また，CVaR は「VaR を超えて発生する損失 $|(\boldsymbol{w}^\top \boldsymbol{x} + b) - y|$ の期待値」（図 5）として定義されるリスク指標であり，以降では $\phi_\beta(\boldsymbol{w}, b)$ で表わす．Rockafellar-Uryasev [8] の結果を適用することで，CVaR を最小にする解 $(\boldsymbol{w}^*, b^*)$ は次の問題を解くことにより得られる．

$$\begin{aligned}\min_{\boldsymbol{w}, b}\ & \phi_\beta(\boldsymbol{w}, b) \\ &= \min_{\boldsymbol{w}, b, \alpha}\ \alpha + \frac{1}{(1-\beta)m} \\ &\quad \times \sum_{i \in M} [|(\boldsymbol{w}^\top \boldsymbol{x}_i + b) - y_i| - \alpha]^+\end{aligned}$$

$[\cdot]^+$ は，ϵ 許容損失の定義で用いられたように，$[a]^+ =$

$\max\{0, a\}$ と定義される．この CVaR 最小化問題の変数名 α を ϵ に変え，パラメータ $(1-\beta)$ を ν に変え，さらに正則化項を加えると，Schölkopf ら [9] によって提示された ν-SVR：

$$\min_{\boldsymbol{w},b,\epsilon} C\left(\nu\epsilon + \frac{1}{m}\sum_{i\in M}[|(\boldsymbol{w}^\top \boldsymbol{x}_i + b) - y_i| - \epsilon]^+\right) + \frac{1}{2}\|\boldsymbol{w}\|_2^2$$

に一致する．ここでは，ν と C が入力パラメータ，ϵ は変数であることに注意したい．この ν-SVR は，最適解 ϵ のもとでの ϵ 許容損失に ℓ_2 正則化項を加えたモデルであり，パラメータ C と C' を適切に対応させれば，次のノルム制約のついた CVaR 最小化問題と等価になる．

$$\min_{\boldsymbol{w},b} \phi_{1-\nu}(\boldsymbol{w}, b) \text{ s.t. } \|\boldsymbol{w}\|_2 \le C' \tag{1}$$

Schölkopf ら [9] は，パラメータ ν が「ϵ を超える損失を持つデータ数の割合」つまり，$|\{i \in M : |h(\boldsymbol{x}_i) - y_i| > \epsilon\}|/m$ の上限値を与えることを示しているが，目的関数そのものの解釈を与えていない．しかし，損失分布に対する CVaR という観点から見れば，損失分布の上位 100ν%の期待値がなるべく小さくなるように $\boldsymbol{w}, b$ を決めようとしていることがわかる．

Gotoh-Takeda [6] は，CVaR を用いて，ν-SVR のモデルの妥当性を示した．ここでは，[3] に従い，閾値 $\theta > 0$ を導入し，絶対損失が θ より大きくなる確率を次のように定義する．

$$P\{(\boldsymbol{x}, y) : |(\boldsymbol{w}^\top \boldsymbol{x} + b) - y| > \theta\}$$

下記の定理はこの確率の上界値を与えている．

定理 2.1 ([6])**.** データ $\boldsymbol{x}$ が半径 B の超球に含まれると仮定する．$\phi_{1-\nu}(\boldsymbol{w}, b) < \theta, \|\boldsymbol{w}\|_2 \le C$ を満たす任意の $(\boldsymbol{w}, b)$ に対して，少なくとも $1-\delta$ の確率で次の不等式が成り立つ．

$$P\{(\boldsymbol{x}, y) : |(\boldsymbol{w}^\top \boldsymbol{x} + b) - y| > \theta\} \le \nu + 2\sqrt{\frac{2}{m}\left(\frac{\kappa}{(\theta - \phi_{1-\nu}(\boldsymbol{w}, b))^2}\log_2(2m) - 1 + \ln\left(\frac{2}{\delta}\right)\right)}$$

ただし，$\kappa := 4c^2(B+\theta+1)^2\{C^2 + B^2(C^2+1) + 1\}$，$c$ はある正の定数．

(1) 式で表される ν-SVR の最適解は「絶対損失が θ より大きくなる確率」の上界値（上式の右辺）を最小化することを意味しており，この定理は ν-SVR のモデルの妥当性を示唆している．

3. ポートフォリオ選択の基本戦略

H. Markowitz (1952) に始まるポートフォリオ選択（portfolio selection）は，資産の収益率分布を所与として，投資家が望むリスク/リターンを実現するような資産配分を求める技術である．具体的には，n 種の投資対象資産（例：株式，債券，派生商品，実物資産）への手持ちの資金の投資配分比率（ポートフォリオ）$\boldsymbol{\pi} := (\pi_1, ..., \pi_n)^\top$ を決める問題である．ここで重要なのは，どういう基準でこれを決めるかである．説明の都合上，トラッキング・ポートフォリオとして知られる方法を例に説明しよう．

3.1 従来のトラッキング・ポートフォリオモデル

この方法は，ベンチマーク[4]の収益率からの乖離（トラッキング・エラー）を最小化するようなポートフォリオを目指す方法である．ここで，n 個の資産を扱うことを想定し，時点 t で観測された収益率データを $\boldsymbol{R}_t := (R_{1,t}, \ldots, R_{n,t})^\top$ と表わし，また，時点 t でのベンチマークの値を I_t と表わすことにする．これらのデータ集合 $(\boldsymbol{R}_t, I_t)$, $t = 1, \ldots, T$，に基づいて，各資産にどれだけの割合 π_i, $i = 1, \ldots, n$，を配分すればよいかを決定するための最適化モデルを考える．ここで決定変数 $\boldsymbol{\pi}$ が満たすべき制約として，予算制約に相当する $\sum_{i=1}^n \pi_i = 1$ が課される．

通常，トラッキング・ポートフォリオの構築は，制約付きの経験損失の最小化として定式化される．具体的には，“ベンチマークとポートフォリオのパフォーマンスの乖離を示す指標” は，ℓ_δ ノルムを用いて，

$$g_\delta(\boldsymbol{\pi}) := \frac{1}{T}\|\boldsymbol{R}\boldsymbol{\pi} - \boldsymbol{I}\|_\delta^\delta$$

と定義することが多い．ここでは，$\boldsymbol{I} = (I_1, \ldots, I_T)^\top$，$\boldsymbol{R}$ は $\boldsymbol{R}_1^\top, \ldots, \boldsymbol{R}_T^\top$ を行ベクトルとして持つ，サイズ $T \times n$ の行列である．

決定変数に対する制約として $\sum_{i=1}^n \pi_i = 1$ 以外のものが課されることもあり，ここではそれを Π で表わすことにする．例えば $\Pi = \{\boldsymbol{\pi} \in \mathbb{R}^n : \boldsymbol{\pi} \ge \boldsymbol{0}\}$ とすれば，空売り（当該銘柄を保有していない状況でその銘柄を売ること）を禁止した制約を表現したことになる．すると，二乗損失を用いたトラッキング・ポートフォリオモデルは，次のように定式化される．

[4] 日経平均株価や TOPIX，S&P500 などに代表される平均株価指数など．

$$\min_{\boldsymbol{\pi}\in\Pi} \frac{1}{T}\|\boldsymbol{R}\boldsymbol{\pi}-\boldsymbol{I}\|_2^2 \ \text{ s.t. } \ \sum_{i=1}^{n}\pi_i = 1 \qquad (2)$$

もし Π として特に制約を設けなければ，(2) は等式制約 1 本の凸 2 次計画問題であり，（目的変数のヘッセ行列が正則であれば）ラグランジュの未定乗数法で簡単に解析解が求まる．また，制約を設けたとしても，Π が凸集合であれば，凸二次計画問題としてさまざまな効率的な手法を用いて解くことができる．また，Π が凸多面体であれば，$g_2(\boldsymbol{\pi})$ の代わりに $g_1(\boldsymbol{\pi})$ を用いると，(2) は線形計画問題として定式化することができるため，そういった利点から，$g_1(\boldsymbol{\pi})$ もまたしばしば用いられる．

標準的な Markowitz のポートフォリオモデルも，(2) の特別なケースとして見なすことができる．Markowitz の平均・分散モデル [7] はポートフォリオの収益率の平均値と標準偏差（バラツキ）が投資家にとって重要であると想定し，一定の平均収益率の見返りに可能な限り収益のバラツキを最小化するようなポートフォリオ選択モデルである．次のように定式化される．

$$\min_{\boldsymbol{\pi}\in\Pi} \ \boldsymbol{\pi}^\top\Sigma\boldsymbol{\pi} \ \text{ s.t. } \ \boldsymbol{r}^\top\boldsymbol{\pi} = \gamma,\ \sum_{i=1}^{n}\pi_i = 1 \qquad (3)$$

$\gamma\,(>0)$ は平均収益率であり，$\boldsymbol{r}\in\mathbb{R}^n$ は資産 i の T 期間の標本平均収益率を第 i 成分に持つようなベクトル，$\Sigma\in\mathbb{R}^{n\times n}$ は資産 i，j の標本共分散を (i,j) 成分にもつ行列である．この目的関数 $\boldsymbol{\pi}^\top\Sigma\boldsymbol{\pi}$ を，すべての成分が 1 の n 次元ベクトル $\boldsymbol{e}_n$ を用いて $\frac{1}{T}\|\boldsymbol{R}\boldsymbol{\pi}-\gamma\boldsymbol{e}_T\|_2^2$ と変形することができるため，(3) をトラッキング・ポートフォリオモデルの特別なケース，つまり，$\boldsymbol{I}:=\gamma\boldsymbol{e}_T$ に設定したケースとして見なすことができる．

3.2 機械学習→金融工学

トラッキング・ポートフォリオモデル (2) について，$\boldsymbol{\pi}\in\Pi$ と $\sum_{i=1}^{n}\pi_i = 1$ の制約のついた回帰モデルとして見なすことができる．そこで，2.1 節で記したように，損失関数を変える，もしくは正則化項を加えることで，将来の変動にうまくあった，つまり予測能力の高いポートフォリオが得られることが期待できる．実際に，既存のトラッキング・ポートフォリオモデルや Markowitz モデルの事後パフォーマンスを向上させるため，近年，正則化を考慮したポートフォリオ最適化モデルがいくつか提案されている．

3.2.1 Brodie らによるモデル

例えば，Brodie ら [1] は Markowitz の平均・分散モデルやトラッキング・ポートフォリオモデルに対して，「ℓ_1 ノルムによる正則化項 $\|\boldsymbol{\pi}\|_1$ が，あるパラメータ値 $C_1>0$ 以下になる」ように制約を加えたモデルを提案した．

$$\min_{\boldsymbol{\pi}\in\Pi} \frac{1}{T}\|\boldsymbol{R}\boldsymbol{\pi}-\boldsymbol{I}\|_2^2 \ \text{ s.t. } \ \sum_{i=1}^{n}\pi_i = 1,\ \|\boldsymbol{\pi}\|_1 \le C_1 \qquad (4)$$

ここで，$C_1\ge 1$ を満たすパラメータ C_1 を想定する（この条件を満たさなければ，(4) は実行可能解を持たないことに注意）．また，$C_1=1$ の場合には，(4) の実行可能領域は $\{\boldsymbol{\pi}\in\mathbb{R}^n : \boldsymbol{\pi}\ge\boldsymbol{0}\}$ となる．これは，空売りを禁止した実行可能領域に一致しており，この正則化項は空売り禁止制約と同じ効果をもたらす．また，等式制約 $\sum_{i=1}^{n}\pi_i=1$ を用いて，$\|\boldsymbol{\pi}\|_1$ を次のように変形できるため，ℓ_1 正則化項は空売りの量を制限しているとも解釈できる．

$$\|\boldsymbol{\pi}\|_1 = 2\sum_{i:\pi_i<0}|\pi_i| + 1$$

(4) は制約付きの lasso 回帰モデルであり，ℓ_1 ノルム制約により最適解が疎（最適解 $\boldsymbol{\pi}^*$ の要素にゼロが多く含まれている状態）になりやすくなる．疎な $\boldsymbol{\pi}^*$ は保有資産銘柄数が少ないことを意味する．投資家にとって，取引手数料は無視できないコストであり，疎なポートフォリオは取引手数料を小さく押さえるという意味では好ましいといえる．

3.2.2 DeMiguel らによるモデル

Markowitz の平均・分散モデル (3) について，平均収益率に関する制約 $\boldsymbol{r}^\top\boldsymbol{\pi}=\gamma$ の $\boldsymbol{r}$ の推定のしにくさが実務家から指摘され，この制約を除いたモデルも扱われることが多い．これは “分散最小化モデル” と呼ばれるモデルである．DeMiguel ら [4] は，分散最小化モデルに対してノルム制約を付加した，ノルム制約付き分散最小化モデル：

$$\min_{\boldsymbol{\pi}\in\Pi} \boldsymbol{\pi}^\top\Sigma\boldsymbol{\pi} \ \text{ s.t. } \ \sum_{i=1}^{n}\pi_i = 1,\ \|\boldsymbol{\pi}\|_\delta \le C_\delta \qquad (5)$$

を提案した．(5) は制約付きの ridge 回帰モデルといえる．

$\delta=2$, $C_2=1/\sqrt{n}$ の場合には，(5) は各資産への配分を等額とした構成比率 $\pi_i=1/n$（$i=1,\ldots,n$）のポートフォリオを唯一の実行可能解，つまり最適解として持つ（これは ℓ_2 ノルム制約と等式制約 $\sum_{i=1}^{n}\pi_i=1$ により明らか）．このポートフォリオ構成法は等配分ポートフォリオと呼ばれ，DeMiguel ら [5] によって，

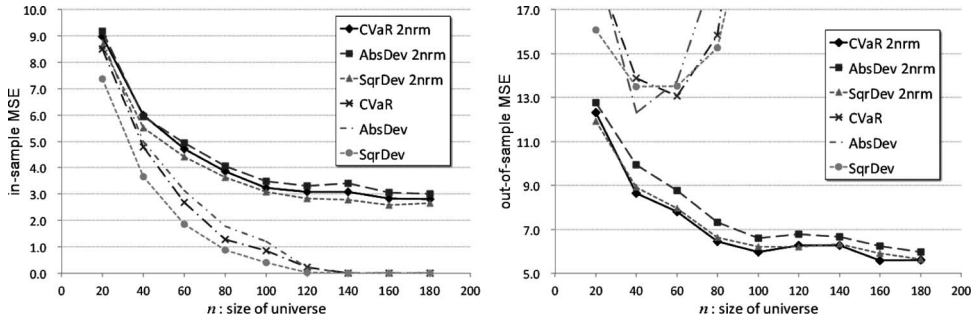

図 6 空売り可能な状況 ($\Pi = \mathbb{R}^n$) における，ポートフォリオモデルの比較（左図の縦軸は事前 MSE，右図の縦軸は事後 MSE．両図とも横軸は投資対象資産数 n）

バラツキは大きいが平均的にみればよい収益をもたらすことが実験的に示されている．

3.2.3 他タイプの損失関数・正則化によるモデル

また，Gotoh-Takeda [6] は (1) 式の ν-SVR をまねて，CVaR 最小化に ℓ_2 正則化を組み合わせた定式化：

$$\begin{aligned} &\min_{\boldsymbol{\pi}\in\Pi,\alpha} \alpha + \frac{1}{(1-\beta)T}\sum_{t=1}^{T}[|\boldsymbol{R}_t^\top\boldsymbol{\pi} - I_t| - \alpha]^+ \\ &\text{s.t.} \sum_{i=1}^{n}\pi_i = 1,\ \|\boldsymbol{\pi}\|_2 \le C_2 \end{aligned} \tag{6}$$

を提案した．このモデルに対し，定理 2.1 が成り立つ．

また，二乗損失に対して ℓ_1 や ℓ_2 正則化項を取り入れるだけでなく，より疎なポートフォリオを得るために ℓ_0 ノルムを用いたモデルが提案されている [2, 10]．また，Yen-Yen [12] は，ℓ_1 ノルムと ℓ_2 ノルムを組み合わせた elastic net [13] を正則化項として用いたモデルを提案している．

3.2.4 正則化項の効果

正則化項の効果を示すため，3 つの損失関数（絶対損失，二乗損失，ϵ 許容損失）それぞれを最小化するモデルと，それらに ℓ_2 正則化項を加えたモデルを比較する．図 6 の “CVaR 2nrm” は式 (6)，“SqrDev 2nrm” は式 (4) のノルム制約を $\|\{\boldsymbol{\pi}\}\|_2 \le C_2$ に変えたもの，“AbsDev 2nrm” はさらに目的関数を $\frac{1}{T}\|\boldsymbol{R}\boldsymbol{\pi} - \boldsymbol{I}\|_1$ に変えたものに対応している．1987 年 5 月〜2009 年 10 月の月次データ（256 月分）のうち，$T = 120$ のデータ（例えば，$\{(\boldsymbol{R}_t, I_t) : t = 1, \ldots, T\}$）を使って学習モデルを構築する．最適解 $\boldsymbol{\pi}^*$ の下での損失 $\frac{1}{T}\|\boldsymbol{R}\boldsymbol{\pi}^* - \boldsymbol{I}\|_2^2$ を訓練誤差と呼ぶ．また，モデル構築に用いていない新たなデータ（例えば，$(\boldsymbol{R}_{t+1}, I_{t+1})$）を用いて評価された誤差 $|\boldsymbol{R}_{t+1}^\top\boldsymbol{\pi}^* - I_{t+1}|^2$ をテスト誤差と呼ぶ．データセットを 1 期ずつずらしてモデルを構築するとともに最適解を求めていくため，136(= 256 − 120) 回，訓練誤差とテスト誤差が計算される．それらについて平均をとったものをそれぞれ，事前 (in-sample)MSE，事後 (out-of-sample)MSE と呼ぶ．正則化項を除いたモデルの方が訓練誤差が小さいのは，モデルの定式化より明らかである．一方，正則化したモデルのほうが事後 MSE が小さいことが見てとれる．つまり，正則化したモデルのほうがしないものよりも予測精度が高いことが確認できる．

また，紙面の都合上割愛するが，正則化項は “収益率データ $\boldsymbol{R}_t, t = 1, \ldots, T$，の不確実性を考慮してロバスト化した経験損失最小化モデル” から導くこともできる [6, 11]．

4. おわりに

金融の用語で「裁定戦略」という言葉がある．これは同じ価値をもつものが 2 つの市場で異なる価格をつけているとすれば，低い評価をつけている市場で買い，高い評価をつけている市場で売り，鞘を稼ぐ操作を指す．

本稿では，筆者らの研究成果を中心に，異分野間の既存の知識をあわせることにより得られた “ちょっと新しいこと” を紹介した．従来 OR が専ら扱ってきた対象以外にも目を向けてみると，新たな裁定戦略の可能性が存在するかもしれない．

参考文献

[1] J. Brodie, I. Daubechies, C. De Mol, D. Giannone and I. Loris, “Sparse and Stable Markowitz Portfolios,” *PNAS*, **106**, 12267–12272 (2009).
[2] T. J. Chang, N. Meade, J. E. Beasley and Y. M. Sharaiha, “Heuristics for Cardinality Constrained Portfolio Optimisation,” *Computers & Operations Research*, **27**, 1271–1302 (2000).
[3] N. Cristianini and J. Shawe-Taylor, “An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods,” Cambridge University Press, Cambridge (2000).

[4] V. DeMiguel, L. Garlappi, F. J. Nogales and R. Uppal, "A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms," *Management Science*, **55**, 798–812 (2009).
[5] V. DeMiguel, L. Garlappi and R. Uppal, "Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?," *Review of Financial Studies*, **22**, 1915–1953 (2009).
[6] J. Gotoh and A. Takeda, "On the Role of Norm Constraints in Portfolio Selection," *Computational Management Science*, **8**, 323–353 (2011).
[7] H. Markowitz, "Portfolio Selection," *Journal of Finance*, **7**, 77–91 (1952).
[8] R. T. Rockafellar and S. Uryasev, "Conditional Value-at-Risk for General Loss Distributions," *Journal of Banking & Finance*, **26**, 1443–1472 (2002).
[9] B. Schölkopf, A. Smola, R. Williamson and P. Bartlett, "New Support Vector Algorithms," *Neural Computation*, **12**, 1207–1245 (2000).
[10] M. Woodside-Oriakhi, C. Lucas and J. E. Beasley, "Heuristic Algorithms for the Cardinality Constrained Efficient Frontier," *European Journal of Operational Research*, **213**, 538–550 (2011).
[11] H. Xu, C. Caramanis and S. Mannor, "Robustness and Regularization of Support Vector Machines," *Journal of Machine Learning Research*, **10**, 1485–1510 (2009).
[12] Y. Yen and T. Yen, "Solving Norm Constrained Portfolio Optimizations via Coordinate-Wise Descent Algorithms," *London School of Economics and Political Science*, UK., http://personal.lse.ac.uk/yen/sp_090111.pdf (2011).
[13] Zou and Hastie, "Regularization and Variable Selection via the Elastic Net," *Journal of Royal Statistical Society: Series B*, **67**, 301–320 (2005).