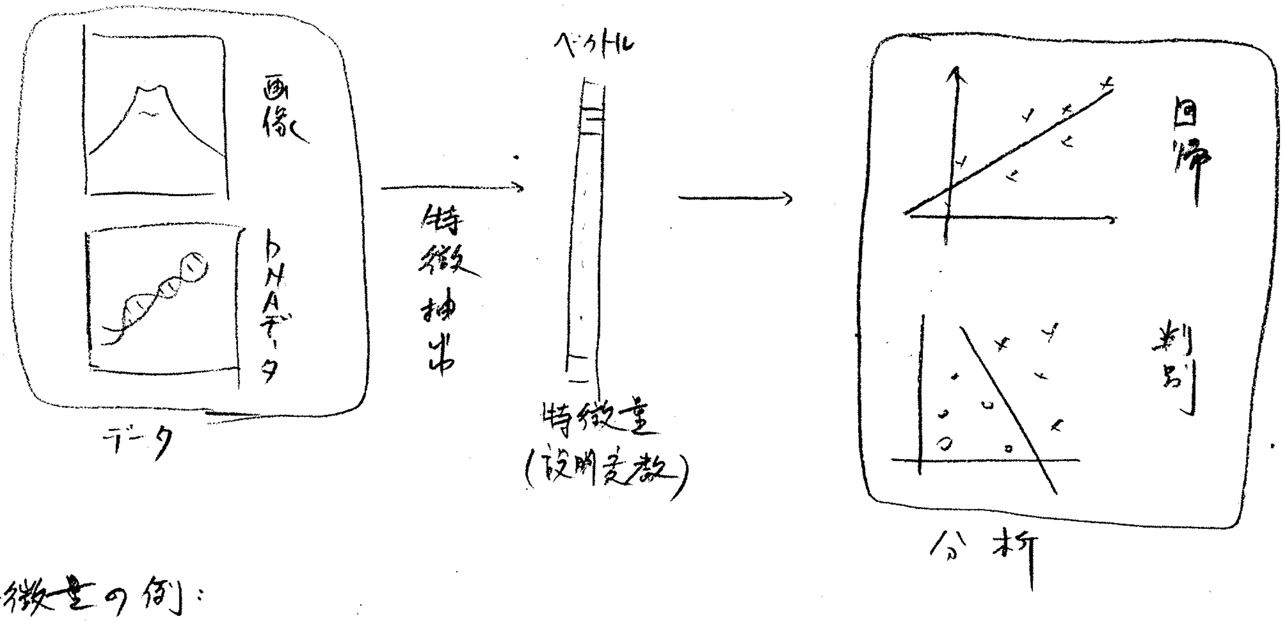
○講義プラン

− 教師あり学習の問題設定




特徴量の例:

画像データ: SIFT, SURF, HoG, Bag of words

バイオインフォ: マイクロアレイデータ, mRNA発現量

特徴抽出後は統計の問題:

$$x \xrightarrow{\text{予測}} y$$

特徴ベクトル　　　　　　ラベル

やりたいこと: $y$を予測する関数$f(x)$をデータから学習したい。

例: 最小2乗法

$$\min_{\beta\in\mathbb{R}^p} \sum_{i=1}^{n} (y_i - x_i^T\beta)^2$$

− ロス関数 (損失関数)

予測値 $f(x)$ がどれだけ $y$ と近いか.

$\ell(y, f(x))$

※ロス関数は目的に合ったものを選ぶ

○回帰: $y \in \mathbb{R}$

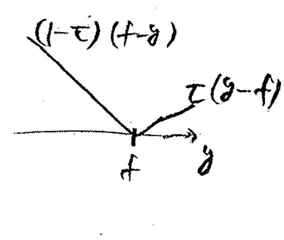
− 2乗ロス

$$\ell(y, f) = (y-f)^2$$

− τ分位ロス

$$\ell(y, f) = (1-\tau)\max(f-y, 0) + \tau \max(y-f, 0)$$

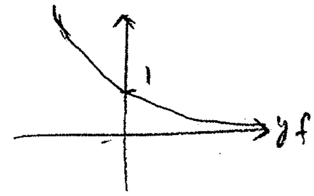
(分位点回帰に用いる)



— 判別: $y \in \{\pm 1\}$

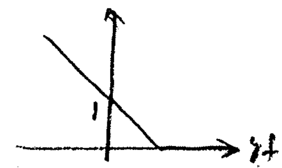
∘ ロジスティックロス

$$l(y,f) = \log(1+\exp(-yf))$$



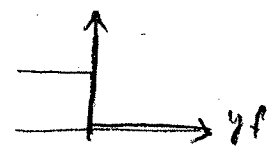

∘ ヒンジロス

$$l(y,f) = \max(1-yf, 0)$$




∘ 0-1ロス

$$l(y,f) = \mathbb{1}[y \neq \mathrm{sign}(f)]$$




— 経験誤差と汎化誤差

n個のサンプル $\{(x_i, y_i)\}_{i=1}^{n}$ が与えられているとする。

データをよく説明する関数を求めたい。

任意の可測関数から選んでくるのは適当でない。(複雑すぎる)

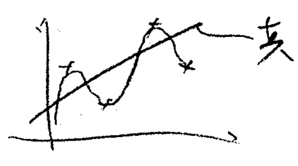
ある適切な関数のクラス $\mathcal{F}$ (モデル)を用意

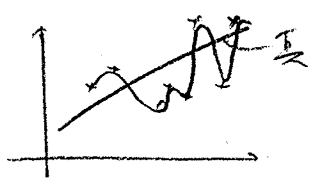
例: $\mathcal{F} = \{f(x) = x^T\beta \mid \beta \in \mathbb{R}^p\}$ (線形モデル)

∘ 経験誤差最小化

$$\hat{f} = \underset{f \in \mathcal{F}}{\arg\min} \sum_{i=1}^{n} l(y_i, f(x_i)) \quad \left( \overset{\text{argmin}}{=:} \hat{L}(f) \right)$$

※ $\mathcal{F}$ が複雑な場合、「過学習」をすることがある。




サンプルを増やしていっても収束しない。

∘ 汎化誤差

$$L(f) = E_{X,Y}[l(Y, f(X))]$$

"未知のデータ(未来のデータ)への当てはまりの良さ。"

本当はこれを良くしたい。

$\boxed{L(f) \text{ と } \hat{L}(f) \text{ には隔たりがある}}$ ← これが問題

今回紹介する方法

正則化学習法

$$\min_{f\in\mathcal{F}} \sum_{i=1}^{n} \ell(y_i, f(x_i)) + \underbrace{R(f)}_{}$$

fの"複雑さ"に対するペナルティ
過学習を防ぐ.

∘ $L_1$-正則化

$$\min_{\beta\in\mathbb{R}^p} \sum_{i=1}^{n} \ell(y_i, x_i^T\beta) + \lambda\|\beta\|_1$$

ただし、$\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|$

∘ カーネル法

$$\min_{f\in\mathcal{H}} \sum_{i=1}^{n} \ell(y_i, f(x_i)) + \lambda\|f\|_{\mathcal{H}}^2$$

ただし、$\mathcal{H}$は"再生核ヒルベルト空間".

---

$L_1$-正則化，スパース学習

- 変数選択

線形モデル（ガウス-マルコフモデル）

$$y_i = x_i^T\beta^* + \varepsilon_i \quad (i=1,\cdots,n)$$

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix},$$ $\beta^*\in\mathbb{R}^p$：真の係数.

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$ ：ノイズ，$\varepsilon_i \sim N(0,\sigma^2)$ i.i.d.

後でサブガウシアンに緩和

$$\Leftrightarrow Y = X\beta^* + \varepsilon$$

nがpに比べて十分大きくないとする

最小2乗法だとうまく推定はできない：$\|\hat{\beta}-\beta^*\|_2^2 = O_p\left(\frac{p}{n}\right)$.

今、$\beta^*$がスパースだとする。

つまり　$k := |\{j \mid \beta_j^* \neq 0\}| \ll p$　とする。

⇒ ムダな情報を削りたい ⇒ 変数選択

Mallow's $C_p$:

$$\min_{\beta} \underbrace{\|Y-X\beta\|^2}_{\text{データへの当てはまり}} + 2\sigma^2 \underbrace{\|\beta\|_0}_{\text{モデルの複雑さ}}$$

ただし、$\|\beta\|_0 = |\{j \mid \beta_j \neq 0\}|$。

⇒ NP-困難 ⇒ 凸緩和

Lasso:
$$\min_{\beta\in\mathbb{R}^p} \|Y-X\beta\|^2 + \lambda\|\beta\|_1$$

※最近は、様々な効率的解法が考案されている。

**命題**

$\|\beta\|_1$ は $\|\beta\|_0$ の $[-1,1]^p$ 上でタイトな凸包絡

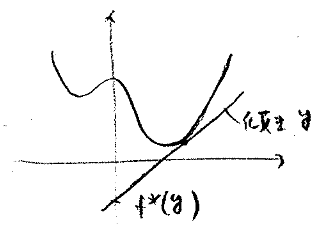
**証明**

(補題)

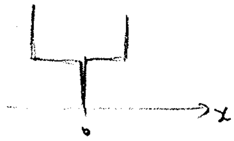
$f^*(y) = \sup_x \{y^T x - f(x)\}$ (Legendre変換)

に対し、

$f^{**}(x)$ は $f(x)$ のタイトな凸包絡




$$f(x) = \begin{cases} \|x\|_0 & (x\in[-1,1]^p) \\ \infty & (\text{otherwise}) \end{cases}$$ に対し、




$$f^{**}(x) = \begin{cases} \|x\|_1 & (x\in[-1,1]^p) \\ \infty & (\text{otherwise}) \end{cases}$$

$[-1,1]^p$ をその外まで延長させることで題意をえる。 //

**実際スパースな推定量が得られる**

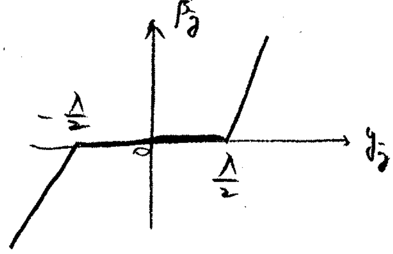
簡単のため、$X = I$ (単位行列) とする。

このとき、$\hat\beta$ のみたすべき条件は

$$2(\hat\beta - Y) + \lambda\, \mathrm{sign}(\hat\beta) \ni 0 \quad (\text{微分して}0\text{とおく})$$

$\mathrm{sign}(\beta) = (\mathrm{sign}(\beta_j))_{j=1}^p$ で、$\mathrm{sign}(\beta_j) = \begin{cases} 1 & (\beta_j > 0) \\ -1 & (\beta_j < 0) \\ [-1,1] & (\beta_j = 0) \end{cases}$ である。

つまり、$\hat\beta_j = \mathrm{sign}(y_j)\left(y_j - \frac{\lambda}{2}\mathrm{sign}(y_j)\right)_+$




Soft-Thresholding 関数

よって、

$|y_j| \leq \frac{\lambda}{2}$ なら $\beta_j = 0$

⟹ スパース解

(原点への縮小)

○ $L_1$-正則化以外のスパース性

- グループ正則化

$$\beta = \begin{bmatrix} \beta_{g_1} \\ \vdots \\ \beta_{g_M} \end{bmatrix}$$

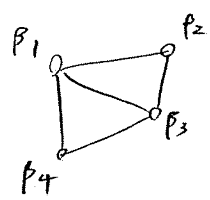
$\beta_{g_i} \in \mathbb{R}^{|g_i|}$

$g_i$ は変数の部分集合

$$\|\beta\|_G := \sum_{j=1}^{M} \|\beta_{g_j}\|_2$$

※グループごと、ごっそり 0 になる。

- 一般化 Fused-Lasso




グラフ $G = (V, E)$

$V = \{1, 2, \ldots, p\}$

が given

$$\|\beta\|_{Fused} = \sum_{(i,j) \in E} |\beta_i - \beta_j|$$

- トレースノルム正則化

$W \in \mathbb{R}^{M \times N}$, 行列

$$\|W\|_{Tr} = \sum_{j=1}^{\min(M,N)} \underbrace{\sigma_j(W)}_{W\text{の}j\text{番目の特異値}}$$

※ $\hat{W}$ は低ランクになりやすい

- 逆共分散選択

- 和型 vs. 畳み込み型

- Lassoの収束レート

Lasso:

$$\hat{\beta} = \arg\min_{\beta} \frac{1}{n} \|Y - X\beta\|_2^2 + \lambda_n \|\beta\|_1$$

〈 $\|\hat{\beta} - \beta^*\|_2^2$ の収束レートを解析〉

仮定

(i) $\max_{i,j} |X_{ij}| \leq 1$

(ii) $J = \{j \mid |\beta_j^*| \neq 0\}$, $|J| =: k$ （$k \ll n$ を想定）

(iii) $\varepsilon_i$ は i.i.d. で、$E[\varepsilon_i] = 0$ かつ、$\exists \sigma > 0$ で

$$E[e^{t\varepsilon_i}] \leq e^{\sigma^2 t^2 / 2} \quad \text{for } {}^\forall t \in \mathbb{R}$$

をみたす。（サブガウシアン）

例：
- $P(\varepsilon_i = 1) = P(\varepsilon_i = -1) = \frac{1}{2}$
- $\varepsilon_i \sim N(0, \sigma^2)$

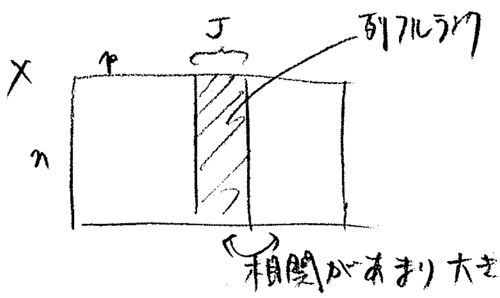
(iv) Restricted Eigenvalue Codition (RE(2k, 3))

$$A := \frac{X^T X}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^T,$$

に対し、

$$\phi_{RE} = \phi_{RE}(2k, 3) = \inf_{\substack{|I| \leq 2k \\ 3\|v_I\|_1 \geq \|v_{I^c}\|_1}} \frac{v^T A v}{\|v_I\|_2^2} \text{ が}$$

$0 < \phi_{RE}$ をみたす。 ほぼスパースという制限




補題（Hoeffdingの不等式）

$\xi_1, \dots, \xi_n$：独立な（同一とは限らない）平均0の確率変数

$E[e^{t\xi_i}] \leq e^{\sigma_i^2 t^2 / 2}$ （${}^\forall t > 0$, $i = 1, \dots, n$）なら

$$P\left( \left| \frac{1}{n} \sum_{i=1}^{n} \xi_i \right| \geq t \right) \leq 2 \exp\left( - \frac{n t^2}{2 \left( \sum_{i=1}^{n} \sigma_i^2 / n \right)} \right)$$

証明

チェビシェフの不等式より、$\tau>0$ に対し

$$P\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i \geq t\right) \leq \frac{E\left[e^{\frac{1}{n}\sum \xi_i \times \tau}\right]}{e^{\tau t}} = \frac{\prod_{i=1}^{n} E\left[e^{\frac{\tau}{n}\xi_i}\right]}{e^{\tau t}}$$

$$\leq \frac{e^{\left(\sum_{i=1}^{n}\sigma_i^2\right)\frac{1}{2}\left(\frac{\tau}{n}\right)^2}}{e^{\tau t}}$$

ここで $\tau = \dfrac{n^2 t}{\sum_{i=1}^{n}\sigma_i^2}$ とおけば題意を得る。

//

補題

仮定(i)と仮定(iii)のもと、$0<{}^{\forall}\delta<1$ に対し

$$P\left(\left\|\frac{1}{n}\varepsilon^T X\right\|_\infty \geq \sigma\sqrt{\frac{2\log(2p/\delta)}{n}}\right) \leq \delta$$

証明

Xの$j$列

$$P\left(\left\|\frac{1}{n}\varepsilon^T X\right\|_\infty \geq r\right) = P\left(\max_{1\leq j\leq p}\left|\frac{1}{n}\varepsilon^T X_j\right| \geq r\right)$$

$$= P\left(\bigcup_{1\leq j\leq p} \quad 〃 \quad\right)$$

$$\leq \sum_{j=1}^{p} P\left( \quad 〃 \quad\right)$$

$$\leq p \max_{1\leq j\leq p} P\left( \quad 〃 \quad\right)$$

ここで、$\xi_i = \varepsilon_i X_{ij}$ とおけば $\frac{1}{n}\varepsilon^T X_j = \frac{1}{n}\sum_{i=1}^{n}\xi_i$ である。

今 $|X_{ij}|\leq 1$ の条件と、$\varepsilon_i$ のサブガウシアン性より、

$$E\left[e^{t\xi_i}\right] \leq e^{\sigma^2 t^2/2} \qquad ({}^{\forall}t>0)$$

である。よって前の補題より

$$P\left(\left|\frac{1}{n}\varepsilon^T X_j\right| > r\right) \leq 2\exp\left(-\frac{n r^2}{2\sigma^2}\right)$$

である。あとは $r = \sigma\sqrt{\dfrac{2\log(2p/\delta)}{n}}$ とすれば題意を得る。

//

## 定理

$0<\delta<1$ を固定する。正則化パラメータを $\lambda_n = 4\sigma\sqrt{\dfrac{2\log(2p/\delta)}{n}}$ とする。

"仮定"のもと、ある定数 $C>0$ が存在し、

$$\|\hat\beta-\beta\|_2^2 \leq C\,\frac{\sigma^2}{\phi_{RE}^2}\times\frac{k\log(2p/\delta)}{n}$$

が確率 $1-\delta$ でなりたつ。

※ $\frac{k\log(p)}{n}$ は $\frac{p}{n}$ より小さくなりません。

## 証明

$\hat\beta$ の定義より、

$$\frac{1}{n}\|Y-X\hat\beta\|_2^2 + \lambda_n\|\hat\beta\|_1 \leq \frac{1}{n}\|Y-X\beta^*\|_2^2 + \lambda_n\|\beta^*\|_1$$

$$\Rightarrow \frac{1}{n}\|X\hat\beta-X\beta^*\|_2^2 + \lambda_n\|\hat\beta\|_1 \leq \frac{2}{n}\varepsilon^TX(\hat\beta-\beta^*) + \lambda_n\|\beta^*\|_1 \quad (\because Y=X\beta^*+\varepsilon)$$

$$\left(\leq \frac{2}{n}\sup_{\beta:\|\beta\|_1\leq\|\hat\beta-\beta^*\|_1}\varepsilon^TX\beta\right)$$

$$\leq 2\left\|\frac{1}{n}\varepsilon^TX\right\|_\infty\|\hat\beta-\beta^*\|_1 + \lambda_n\|\beta^*\|_1$$

ここで前の補題より、$P\left(\|\frac{1}{n}\varepsilon^TX\|_\infty \geq \frac{\lambda_n}{4}\right) \leq \delta$ である。

以後、$\|\frac{1}{n}\varepsilon^TX\|_\infty \leq \frac{\lambda_n}{4}$ とする。

$$\Rightarrow \frac{1}{n}\|X\hat\beta-X\beta^*\|_2^2 + \lambda_n\|\hat\beta\|_1 \leq \frac{\lambda_n}{2}\|\hat\beta-\beta^*\|_1 + \lambda_n\|\beta^*\|_1 \quad \text{—— ①}$$

ここで、$J^c$ を $|\hat\beta_j|$ の降順に並べると

$$|\hat\beta_{j_1}| \geq |\hat\beta_{j_2}| \geq \cdots \geq |\hat\beta_{j_{p-|J|}}|.$$

これの上から順に $k$ 個のインデックス $\{j_1,\ldots,j_k\}\subseteq J^c$ をとってくる。

$$I = J\cup\{j_1,\ldots,j_k\}$$

とおく。すると、$\beta^*_j=0\ (\forall j\in J^c)$ に気をつけると、①より

$$\frac{1}{n}\|X\hat\beta-X\beta^*\|_2^2 + \lambda_n\|\hat\beta_{I^c}\|_1 \leq \frac{\lambda_n}{2}\left(\|\hat\beta_I-\beta^*_I\|_1 + \|\hat\beta_{I^c}-\underbrace{\beta^*_{I^c}}_{=0}\|_1\right) + \lambda_n\overbrace{\left(\|\beta^*_I\|_1-\|\hat\beta_I\|_1\right)}^{\leq\|\beta^*_I-\hat\beta_I\|_1}$$

$$\Rightarrow \frac{1}{n}\|X\hat\beta-X\beta^*\|_2^2 + \frac{\lambda_n}{2}\|\hat\beta_{I^c}\|_1 \leq \frac{3}{2}\lambda_n\|\hat\beta_I-\beta^*_I\|_1 \quad \text{—— ②}$$

よって、$\hat\beta-\beta^*$ は

$$\|(\hat\beta-\beta^*)_{I^c}\|_1 \leq 3\|(\hat\beta-\beta^*)_I\|_1$$

をみたす。

これより、$\hat{\beta}-\beta^*$ は仮定の(iv)の条件をみたすので

$$\phi_{RE}\|\hat{\beta}_I-\beta^*_I\|_2^2 \leqq \frac{1}{n}\|X(\hat{\beta}-\beta^*)\|^2$$

である。よって ②より

$$\|\hat{\beta}_I-\beta^*_I\|_2^2 \leqq \frac{1}{\phi_{RE}}\cdot\frac{3}{2}\lambda_n\|\hat{\beta}_I-\beta^*_I\|_1$$

$$\leqq \frac{3}{2\phi_{RE}}\lambda_n\sqrt{2k}\,\|\hat{\beta}_I-\beta^*_I\|_2 \quad (\because |I|=2k)$$

$$\Longrightarrow \|\hat{\beta}_I-\beta^*_I\|_2 \leqq \frac{3}{2\phi_{RE}}\sqrt{2k}\,\lambda_n$$

また、

$$\|\hat{\beta}-\beta^*\|_2-\|\hat{\beta}_I-\beta^*_I\|_2 \leqq \|\hat{\beta}_{I^c}-\beta^*_{I^c}\|_2$$

$$\leqq \sqrt{\|\hat{\beta}_{I^c}-\beta^*_{I^c}\|_\infty\|\hat{\beta}_{I^c}-\beta^*_{I^c}\|_1}$$

$$\leqq \sqrt{\frac{\|\hat{\beta}_I-\beta^*_I\|_1}{k}\|\hat{\beta}_{I^c}-\beta^*_{I^c}\|_1}$$

↑ $I$の作り方より

$$\leqq \sqrt{\frac{3}{k}}\,\|\hat{\beta}_I-\beta^*_I\|_1$$

$$\leqq \sqrt{6}\,\|\hat{\beta}_I-\beta^*_I\|_2$$

以上より

$$\|\hat{\beta}-\beta^*\|_2 \leqq (1+\sqrt{6})\|\hat{\beta}_I-\beta^*_I\|_2$$

$$\leqq (1+\sqrt{6})\times\frac{3}{\sqrt{2}\phi_{RE}}\cdot\sqrt{k}\,\lambda_n$$

//

- REC条件も満たすXの構成

Xが性質の良いランダム行列の場合、REC条件は高い確率で成り立つ。

RI条件 (Restricted Isometry 条件, $RI(d,\delta)$)

$d$-スパースな全ての $\beta \in \mathbb{R}^p$ に対し ($\|\beta\|_0 \leq d$)

$$(1-\delta)\|\beta\|_2 \leq \frac{\|X\beta\|_2}{\sqrt{n}} \leq (1+\delta)\|\beta\|_2$$

が成り立つ。

補題

$1/5 > \delta > 0$, $0 < k < p$, $b > 0$ とし

$$d = k\left\{1 + \left(\frac{16(3b)^2(3b+1)}{\delta^2}\right)\right\}$$

とする。このとき、$X \in \mathbb{R}^{n\times p}$ が

$RI(d,\delta)$ をみたすなら、$RE(k,b)$ をみたし

$$0 < \sqrt{1-5\delta} \leq \phi_{RE}(k,b,X)$$

である。

(証明は省略)

よって、上で定義される $d\ (=d(k,b))$ において、$RI(d,\delta)$ が成り立てば良い。

定義 (isotropic sub-Gaussian)

$\mathbb{R}^p$ に値をとる確率変数 $Y$ が次の条件をみたすとき、
$Y$は isotropic ($\alpha$-)subGaussian であると言う

(1) $^\forall y \in \mathbb{R}^p$ に対し、

$$E\langle Y,y\rangle^2 = \|y\|_2^2 \qquad (\text{等方的})$$

(2) ある $\alpha$ が存在して、$^\forall y \in \mathbb{R}^p$ に対して

$$\|\langle Y,y\rangle\|_{\psi_2} := \inf\{t \mid E[\exp(\langle Y,y\rangle^2/t^2)] \leq 2\} \leq \alpha\|y\|_2$$

である。

2であるべき 正しくは $E[\ ]-1 \leq 1$ の表記

例 (1) $Y_i \sim N(0,1)$ (i.i.d.) のとき、 isotropic subGaussian.

(2) $P(Y_i = 1) = P(Y_i = -1) = \frac{1}{2}$ (i.i.d.) のとき、isotropic sub Gaussian
(see Vershynin)

※ sub-Exponential でも似たような議論が可能

定理

$X = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix} \in \mathbb{R}^{n \times p}$ としたとき、各行ベクトル $x_i \in \mathbb{R}^p$ は独立な isotropic $\alpha$-subGaussian であるとする。なお、$d \le p$, $0 < \delta < 1$ に対し、

$$n \ge \frac{80 d \alpha^4}{\delta^2} \log\left(\frac{12 e p}{d \delta}\right) \qquad \left(n = \Omega\left(d \log\left(\frac{p}{d}\right)\right)\right.$$
$\Omega(p)$ ではない)

ならば、全ての $d$-スパースな $\beta$ に対し

$$(1-\delta)\|\beta\|_2 \le \frac{1}{\sqrt{n}}\|X\beta\|_2 \le (1+\delta)\|\beta\|_2 \qquad (RI(d,\delta))$$

が確率 $1 - 2\exp(-\delta^2 n / 80\alpha^4)$ で成り立つ。

※ Johnson-Lindenstrauss の補題ともよばれている。距離空間の低次元ユークリッド空間への等長埋め込み。

証明

補題

$Y_i \in \mathbb{R}$ $(i=1,\dots,n)$ を独立な確率変数とする。
$E[Y_i^2] = 1$, $\|Y_i\|_{\psi_2} \le \alpha$ が $\forall i = 1,\dots,n$ で成り立つとき
$\forall t \in (0,1)$ において

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n Y_i^2 - 1\right| > t\right) \le 2\exp\left(-\frac{t^2 n}{10\alpha^4}\right)$$

である。 ($t > 1$ ではでもない)

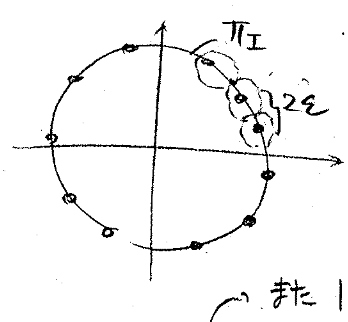
(証明は略)

RI-条件を示すには、$\|\beta\|_2 = 1$ として一般性を失わない。

今、$I \subseteq \{1,2,\dots,p\}$ に対し、$\Pi_I := \{\beta \in \mathbb{R}^p \mid \beta_j = 0\ (j \notin I),\ \|\beta\|_2 = 1\}$ とする。

$\Pi_I$ を有限個の点で代表させることを考える。

今、$H_I \subseteq \Pi_I$ を $\Pi_I$ の部分集合で $\forall \beta \in \Pi_I$ に対し $\exists g \in H_I$ で $\|\beta - g\|_2 \le \varepsilon$ とできるようにもって来る。




$$\binom{p}{d} = \frac{p \cdots (p-d+1)}{d!}$$
ここで $d! = \sqrt{2\pi d}\left(\frac{d}{e}\right)^d e^{\lambda_d}$
ただし $\frac{1}{12d+1} \le \lambda_d \le \frac{1}{12d}$
$\Rightarrow d! \ge \left(\frac{d}{e}\right)^d$

$$|H_I| \le \left(\frac{3}{\varepsilon}\right)^{|I|}$$

とすることができることが知られている。
(超球面のカバリングナンバー)

さらに、

$$\bigcup_{|I|=d} \Pi_I = \Pi, \quad \bigcup_{|I|=d} H_I = H$$

にすれば、$\forall \beta \in \Pi$ に対し $\exists g \in H$ で $\|\beta - g\|_2 \le \varepsilon$ である。

また $|H| \le \left(\frac{3}{\varepsilon}\right)^d \binom{p}{d}$
$\le \left(\frac{3}{\varepsilon}\right)^d \left(\frac{ep}{d}\right)^d$
$= \exp\left(d \log\left(\frac{3ep}{d\varepsilon}\right)\right)$

まずは $H$ に限れば RI条件が成り立つことを示そう。

ある $\beta\in H$ を1つ固定すれば、補題より

$$
\begin{aligned}
&P\left(\left|\frac{1}{n}\|X\beta\|_2^2 - 1\right| \ge t\right)\\
&= P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\langle x_i,\beta\rangle^2 - 1\right| \ge t\right)\\
&\le 2\exp\left(-\frac{nt^2}{10\alpha^4}\right)
\end{aligned}
$$

$(0<t<1)$ （$\because$ $x_i$ は isotropic $\alpha$-subGaussian）

である。よって、

$$
\begin{aligned}
&P\left(\left|\frac{1}{n}\|X\beta\|_2^2 - 1\right| \ge t,\ \forall\beta\in H\right)\\
&\le \sum_{\beta\in H} P\left(\left|\frac{1}{n}\|X\beta\|_2^2 - 1\right| \ge t\right)\\
&\le 2|H|\exp\left(-\frac{nt^2}{10\alpha^4}\right)\\
&\le 2\exp\left(\log(|H|) - \frac{nt^2}{10\alpha^4}\right) \le 2\exp\left(d\log\left(\frac{3ep}{d\varepsilon}\right) - \frac{nt^2}{10\alpha^4}\right)
\end{aligned}
$$

$\left[\ n \ge \dfrac{20d\alpha^4}{t^2}\log\left(\dfrac{3ep}{d\varepsilon}\right)\right.$ とすると、$\left.\right]$

$$
\le 2\exp\left(-\frac{nt^2}{20\alpha^4}\right).
$$

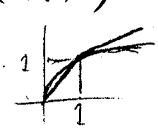
※

よって、$\forall\beta\in H$ に対し、

$$
1-t \le \frac{1}{n}\|X\beta\|_2^2 \le 1+t
$$

$$
\Rightarrow\ 1-t \le \frac{1}{\sqrt{n}}\|X\beta\|_2 \le 1+\frac{t}{2} \qquad (t\in(0,1))
$$

with prob $1-2\exp\left(-\dfrac{nt^2}{20\alpha^4}\right)$.




ここで、$\displaystyle\sup_{\beta\in\Pi}\frac{1}{\sqrt{n}}\|X\beta\|_2 =: \left\|\frac{X}{\sqrt{n}}\right\|_\Pi$ とおく。

$\forall\beta\in\Pi$ に対し、$\exists q\in H$ で $\|\beta-q\|_2\le\varepsilon$ かつ $\dfrac{\beta-q}{\|\beta-q\|_2}\in\Pi$ とできるので、

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\|X\beta\|_2 = \frac{1}{\sqrt{n}}\|X(\beta-q+q)\|_2 &\le \frac{1}{\sqrt{n}}\|X(\beta-q)\|_2 + \frac{1}{\sqrt{n}}\|Xq\|_2\\
&\le \varepsilon\left\|\frac{X}{\sqrt{n}}\right\|_\Pi + 1+\frac{t}{2}
\end{aligned}
$$

よって、左辺で $\sup_\beta$ をとることで

$$
\left\|\frac{X}{\sqrt{n}}\right\|_\Pi \le \frac{1+\frac{t}{2}}{1-\varepsilon}
$$

であることがわかる。同様にして $\dfrac{1}{\sqrt{n}}\|X\beta\|_2 \ge 1-t-\left(\dfrac{1+\frac{t}{2}}{1-\varepsilon}\right)\varepsilon$ がわかる。

これらより，$\varepsilon = \frac{t}{1+2t}$ $\left(0 < t < \frac{1}{2}\right)$ とすることで，

$$1 - 2t \leq \frac{1}{\sqrt{n}} \|X\beta\|_2 \leq 1 + 2t \qquad (\forall \beta \in \Pi)$$

を得る。 $\delta = 2t$ とすることで題意を得る。

//

より詳しくは，

Rudelson & Zhou : Reconstruction From Anisotropic Random Measurements.
IEEE Trans. Info. Th., vol. 59, No. 6, 3434–3447.

を参照のこと。

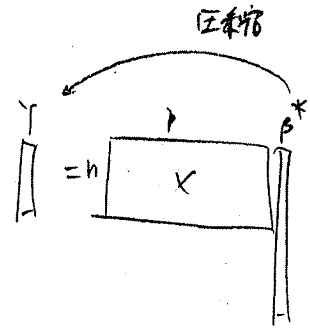
応用例：

- 圧縮センシング

あるβ*に対し，$Y = X\beta^*$ を観測

$X \in \mathbb{R}^{n \times p}$ で $n \ll p$ とする。

⟶ p次元を n次元に圧縮




$$\hat{\beta} = \underset{\beta \in \mathbb{R}^p}{\mathrm{argmin}} \left\{ \|\beta\|_1 \mid X\beta = Y \right\}$$

定理 (Candès, 2008)

$X$ が $RI(\delta, 2k)$ を $\delta < \sqrt{2} - 1$ に対しみたしているとき，
$\forall \beta^*$：k-スパース に対し，

$$\hat{\beta} = \beta^*$$

となる。 (Exact recovery)

$n \geq C k \log(p/k)$ で十分

- 大規模行列の SVD (Halko, Martinsson, Tropp ; 2011)

$A \in \mathbb{R}^{M \times N}$ : 大規模，ほぼ低ランク or 特異値の減少が速い