機械学習アルゴリズムにおける損失関数と不確実性集合の双対性

T. Kanamori (Nagoya Univ.), A. Takeda (UT), T. Suzuki (Titech), and S. Fujiwara

Sep. 2, 2015

統計多様体の幾何学とその周辺 **(**北大**)**

講演の内容

機械学習における２値判別問題

- ２つのアプローチの関連を調べる
  - 損失関数
  - 不確実性集合

  不確実性集合：
  損失関数 $\ell(z)$ の共役 $\ell^*(\alpha)$ の **level set** で表せる

- 不確実性集合アプローチの統計的性質

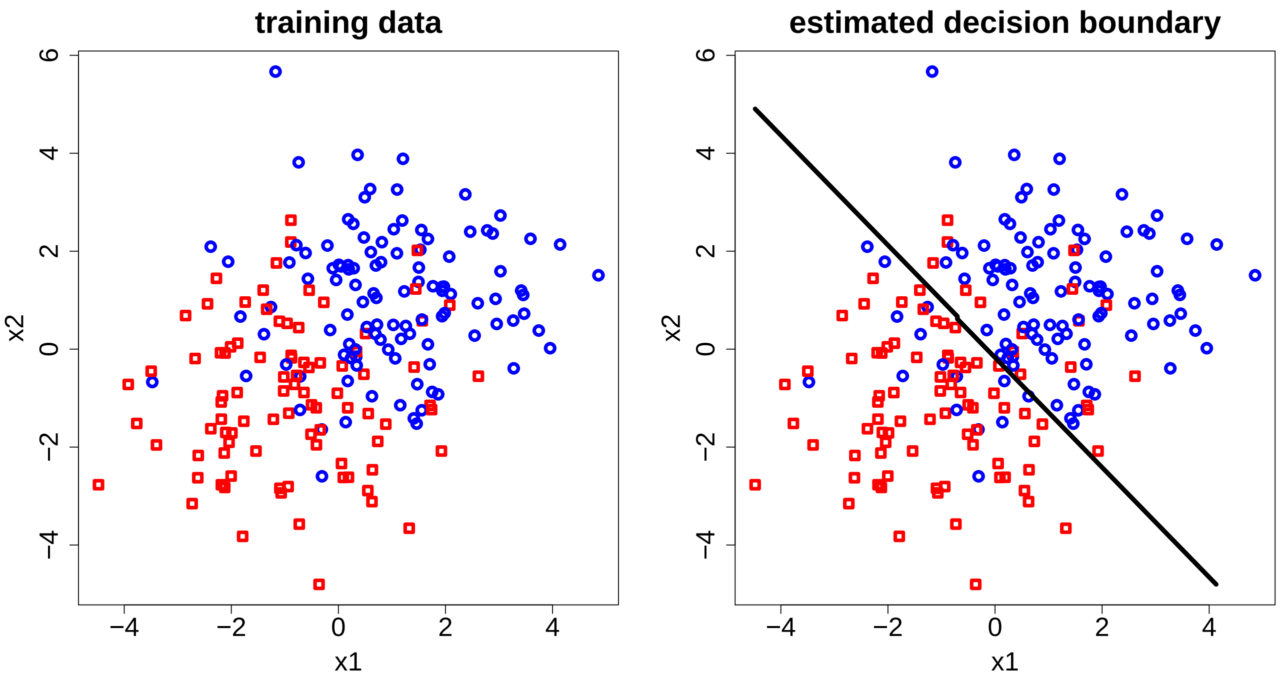
- その後の展開：ロバスト性解析への応用

２値判別

- **i.i.d. データ** $(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_m, y_m) \in \mathcal{X} \times \{+1, -1\}$**.**

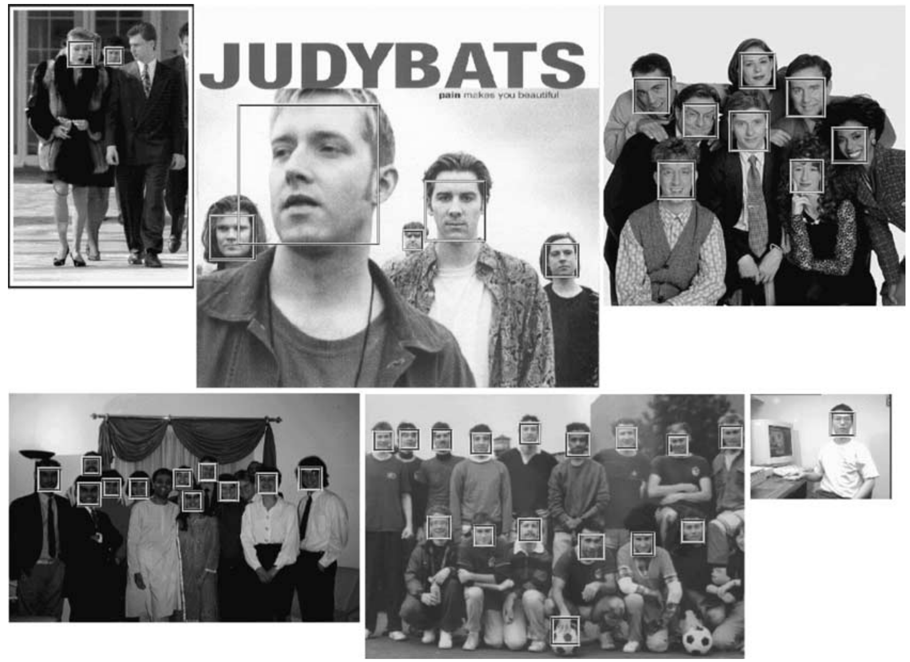
  $\boldsymbol{x}_i$：入力ベクトル，y_i：出力ラベル

- 目標：新たな入力 $\boldsymbol{x}$ のラベル $y \in \{+1, -1\}$ を当てる．



顔検出：顔がある位置を検出．デジカメなどに搭載．

学習データ（正例）



検出結果

P. Viola & M. J. Jones, Robust Real-Time Face Detection Journal International Journal of Computer Vision, 2004

共役性：ルジャンドル変換

凸関数 $\ell : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$

$$\text{ルジャンドル変換：} \quad \ell^*(\alpha) = \sup_{z \in \mathbb{R}} \{\, \alpha z - \ell(z) \,\}.$$

(凸関数を接線の集合で表現**)**

$\ell(z)$ が適当な条件 **(convex, closed, proper)** を満たす

$$\implies (\ell^*)^* = \ell$$

- 本発表では「$\mathbb{R}$ 上の凸関数のルジャンドル変換」を主に扱う.

- 共役性： $\ell(z)$ で書けているものを $\ell^*(\alpha)$ で表したり，その逆を考えたりする．
  - 機械学習アルゴリズムの考察に有効
  - $\ell(z)$：**2**値判別の損失関数．
    ヒンジ損失，２乗損失，指数損失，**etc.**

ラベルの予測

1. データから線形判別関数 $f(x) = \boldsymbol{w}^T \boldsymbol{x} + b$ を推定.

2. $f(\boldsymbol{x})$ の符号でラベルを予測. 理想的には・・・

$$\left.\begin{array}{l} (\boldsymbol{x}, y) = (\boldsymbol{x}, +1) \implies f(\boldsymbol{x}) > 0 \\ (\boldsymbol{x}, y) = (\boldsymbol{x}, -1) \implies f(\boldsymbol{x}) < 0 \end{array}\right\} \iff f(\boldsymbol{x})y > 0$$

- $\boldsymbol{x} = (x_1, x_2)$ だけでなく $\boldsymbol{x} = (x_1, x_2, x_1^2, x_2^2, x_1 x_2)$ などもあり.
- $\boldsymbol{x}$ を関数 $k(\cdot, \boldsymbol{x})$ とするとカーネル法.

予測誤差を測る基準

- $f(x)$ の期待予測誤差：

$$\Pr\{Y \neq \mathrm{sign}(f(X))\} = \mathbb{E}[\ell_0(-yf(\boldsymbol{x}))]$$

0-1損失： $\ell_0(z) = [\![z \geq 0]\!]$.
間違えた個数 $(-y_i f(x_i) \geq 0)$ をカウント

- **Bayes error** $:= \inf_{f:\text{可測}} \Pr\{Y \neq \mathrm{sign}(f(X))\}$ を達成する $f(x)$ が最適.

データ数 $m \to \infty$ のとき：

$$\frac{1}{m}\sum_{i=1}^{m} \ell_0(-y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b)) \xrightarrow{p} \mathbb{E}[\ell_0(-y(\boldsymbol{w}^T\boldsymbol{x} + b))]$$

誤判別のデータ数を最小化

$$\min_{\boldsymbol{w},b} \ \frac{1}{m}\sum_{i=1}^{m} \ell_0(-y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b)) \quad \text{s.t. } \boldsymbol{w} \in \mathbb{R}^d,\ b \in \mathbb{R}$$

期待予測誤差を小さくする推定量が得られる(と期待される)

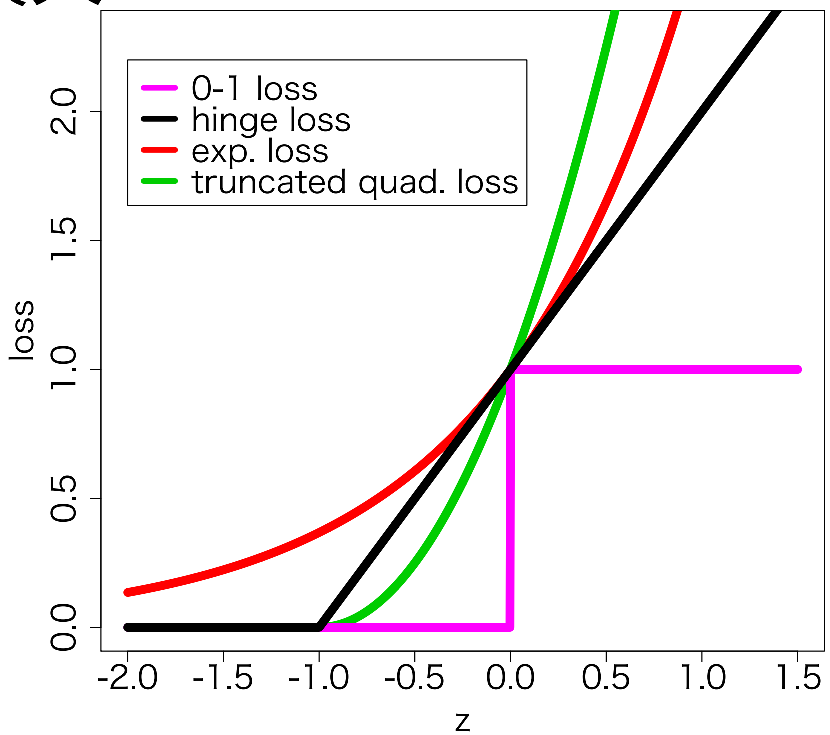
Surrogate loss (代替損失)

- **0-1** 損失 $\ell_0(z) = [\![\, z \geq 0 \,]\!]$ は凸でない：一般に最小化は困難.
- **0-1** 損失 $\ell_0(z)$ の代替損失 $\ell(z)$ を用いる.
  - $\ell_0(z) \leq \ell(z)$ を満たす (適当に定数倍すれば)
  - $\ell(z)$ は凸関数：最小化しやすい

損失関数アプローチ：代替損失を用いる方法

効率的な最適化計算が可能

- $\ell(z) = \max\{1+z,\,0\}$　ヒンジ損失
- $\ell(z) = (\max\{1+z,\,0\})^2$　**(truncated-)** 2乗損失
- $\ell(z) = e^z$　指数損失



代替損失による学習

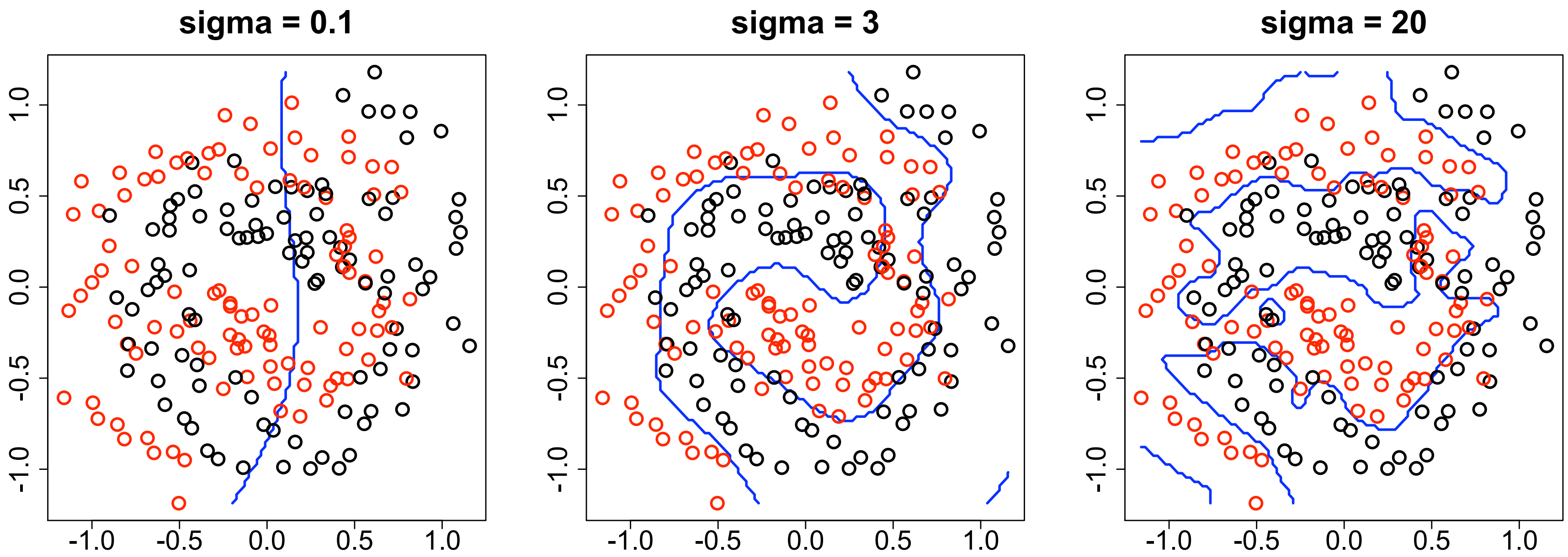
モデル: $f(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + b$, 代替損失：$\ell(z)$

$$\min_{\boldsymbol{w},b} \frac{1}{m}\sum_{i=1}^{m} \ell(-y_i(\boldsymbol{w}^T \boldsymbol{x}_i + b)) \longrightarrow \ \widehat{\boldsymbol{w}}, \widehat{b}$$

- 将来のデータ $\boldsymbol{x}$ のラベルを $\mathrm{sign}(\widehat{\boldsymbol{w}}^T \boldsymbol{x} + \widehat{b})$ で予測

Overfitting と 正則化

データの **complexity** $\ll$ モデルの **complexity**
$\Longrightarrow$ $\widehat{f}(x)$ の予測精度が悪くなる **(overfitting)**



モデルを制約する

(モデルの **complexity** をデータの **complexity** に合わせる**)**

$$\min_{\boldsymbol{w},b} \frac{1}{m}\sum_{i=1}^{m} \ell(-y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b)), \qquad \underline{\|\boldsymbol{w}\|^2 \le \lambda^2}$$

$\lambda\,(\ge 0)$：正則化パラメータ

例：ソフトマージン SVM

ヒンジ損失を使う

$$\min_{\boldsymbol{w},b} \frac{1}{m}\sum_{i=1}^{m}\max\{1-y_i(\boldsymbol{w}^T\boldsymbol{x}_i+b),\,0\}, \quad \|\boldsymbol{w}\|^2 \le \lambda^2$$

さまざまな良い性質

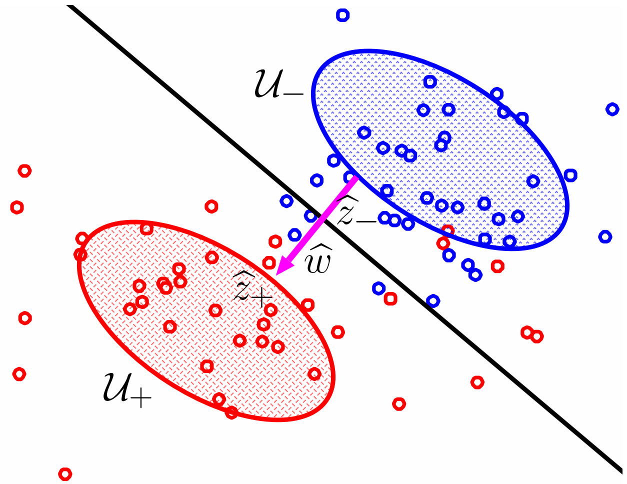
- 凸２次最適化
- カーネル関数(再生核ヒルベルト空間) による柔軟なモデリング
- 統計的性質も優れている.

不確実性集合アプローチ

不確実性集合：ロバスト最適化の分野の用語.

入力 $\boldsymbol{x}$ の空間で幾何的に考える.

- 各ラベルの「不確実性」を $\mathcal{U}_+, \mathcal{U}_- \subset \mathcal{X}$ で表現 ($\mathcal{U}_+ \cap \mathcal{U}_- = \emptyset$).
  「信頼領域」をイメージ.
- $\mathcal{U}_+$, $\mathcal{U}_-$ 間のマージン最大化.



アルゴリズム：不確実性集合アプローチ

1. 各ラベルに対して「データの不確実性」を表す集合を作る

$\Longrightarrow \mathcal{U}_+, \mathcal{U}_- \subset \mathcal{X}$ **(**学習データに依存**)**

2. $\mathcal{U}_+, \mathcal{U}_-$ を最もよく分離する平面を $\widehat{f}(\boldsymbol{x}) = \widehat{\boldsymbol{w}}^T \boldsymbol{x} + \widehat{b}$ とする.

$$\text{最小距離問題：}\quad \min_{\boldsymbol{z}_+, \boldsymbol{z}_-} \|\boldsymbol{z}_+ - \boldsymbol{z}_-\|, \quad \boldsymbol{z}_+ \in \mathcal{U}_+, \quad \boldsymbol{z}_- \in \mathcal{U}_-$$

$\Longrightarrow$ **opt. sol.** $\widehat{\boldsymbol{z}}_+, \widehat{\boldsymbol{z}}_-$

$$\widehat{\boldsymbol{w}} \propto \widehat{\boldsymbol{z}}_+ - \widehat{\boldsymbol{z}}_-, \quad \widehat{b} = -\widehat{\boldsymbol{w}}(\widehat{\boldsymbol{z}}_+ + \widehat{\boldsymbol{z}}_-)/2 \ (\text{など})$$

例：ハードマージンSVM

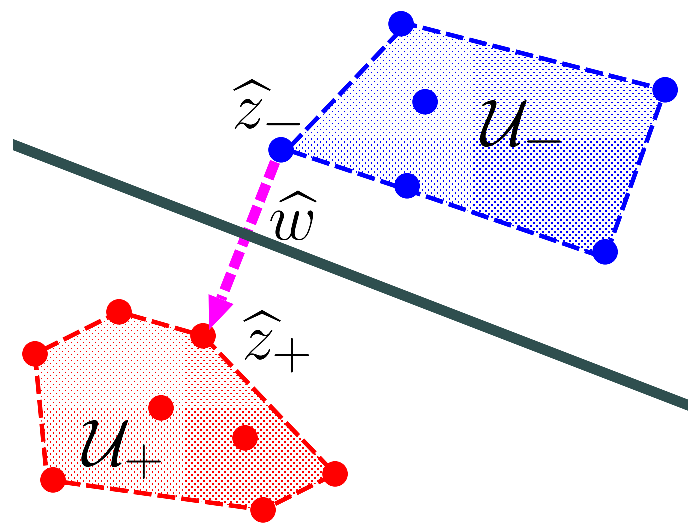
仮定：データは線形分離可能

1. データから不確実性集合を作る

$$\mathcal{U}_+ = \mathrm{conv}\{\boldsymbol{x}_i : y_i = +1\}, \quad \mathcal{U}_- = \mathrm{conv}\{\boldsymbol{x}_i : y_i = -1\}$$

2. マージン最大化：$\mathcal{U}_+$ と $\mathcal{U}_-$ を最も大きく分離

$\longrightarrow \widehat{f}(\boldsymbol{x}) = \widehat{\boldsymbol{w}}^T \boldsymbol{x} + \widehat{b}$



例：Maximum margin MPM [7]

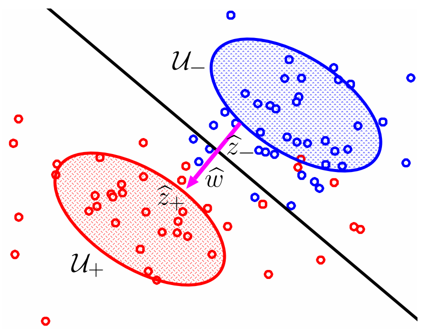
- **max-margin MPM uses ellipsoidal uncertainty sets:**

$$\{\boldsymbol{x}_i : y_i = +1\} \longrightarrow \mathcal{U}_+ = \{\, \widehat{\boldsymbol{\mu}}_+ + \widehat{\Sigma}_+^{1/2}\boldsymbol{u} : \|\boldsymbol{u}\|^2 \le c_+ \,\}$$

$$\{\boldsymbol{x}_i : y_i = -1\} \longrightarrow \mathcal{U}_- = \{\, \widehat{\boldsymbol{\mu}}_- + \widehat{\Sigma}_-^{1/2}\boldsymbol{u} : \|\boldsymbol{u}\|^2 \le c_- \,\}$$

$\widehat{\boldsymbol{\mu}}_\pm, \widehat{\Sigma}_\pm$: **estimated mean and variance-covariance matrices**

- **maximum margin hyperplane between two ellipsoids.**



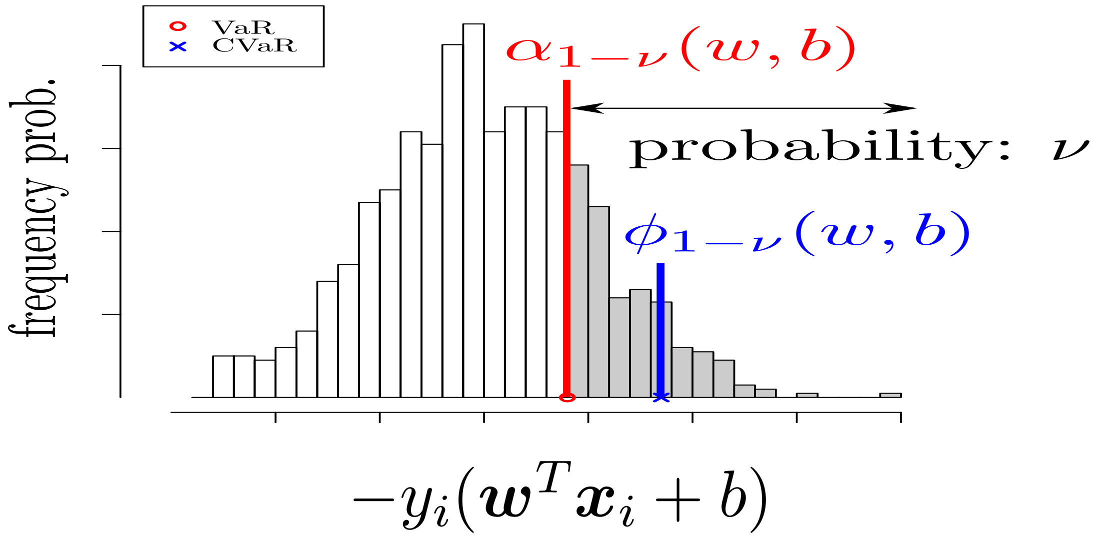
$\min_{\boldsymbol{z}_\pm} \|\boldsymbol{z}_+ - \boldsymbol{z}_-\|^2$ s.t. $\boldsymbol{z}_\pm \in \mathcal{U}_\pm$.

(optimization: SOCP)

ν-SVMにおける損失関数と不確実性集合の関連

$$\nu\text{-}\mathbf{SVM}: \min_{\boldsymbol{w},b,\rho} -\nu\rho + \frac{1}{m}\sum_{i=1}^{m} \max\{\rho \underline{-y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b)},\ 0\} + \frac{1}{2}\|\boldsymbol{w}\|^2$$

- $-y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b)$ の値が(平均的に)小さくなるように$\boldsymbol{w}, b$を選ぶ



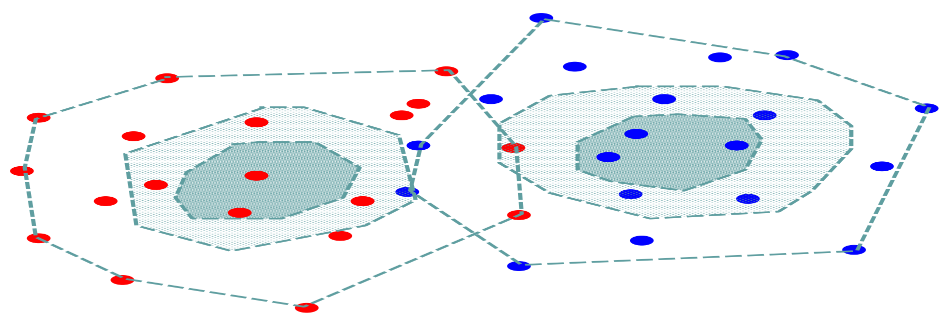
ν-SVM の不確実性集合

- **ν-SVM** の双対問題を導出：

$$\min_{\alpha} \Big\| \sum_{i:y_i=+1} \alpha_i \boldsymbol{x}_i - \sum_{j:y_j=-1} \alpha_j \boldsymbol{x}_j \Big\|^2$$

$$\text{subject to} \quad \sum_{i:y_i=+1} \alpha_i = \sum_{j:y_j=-1} \alpha_j = 1, \quad 0 \le \alpha_i \le \frac{2}{m\nu}$$

- $\mathcal{U}_\pm$：縮小凸包 **(reduced convex-hull)**



ν-SVM の一般化

- ν-**SVM**：ヒンジ損失 $\Longleftrightarrow$ 縮小凸包
- ν-**SVM** の拡張： 損失関数と不確実性集合の関係を一般化
- 一般化する利点：

  (統計的性質がよく分からない**)** 不確実性集合アプローチを

  (統計的性質がよく分かる**)** 損失関数アプローチに帰着

ν-**SVM** :

$$\min_{\boldsymbol{w},b,\rho} -\nu\rho + \frac{1}{m}\sum_{i=1}^{m} \max\{\rho - y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b),\, 0\} + \frac{1}{2}\|\boldsymbol{w}\|^2$$

一般化: $\underline{\ell(z)\text{ を凸，単調非減少，非負}}$とする.

$$\min_{\boldsymbol{w},b,\rho} -2\rho + \frac{1}{m}\sum_{i=1}^{m} \ell(\rho - y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b)) \quad \text{s.t.} \quad \|\boldsymbol{w}\|^2 \le \lambda^2$$

$\ell(z) = \max\{2z/\nu,\, 0\}$ とするとν-**SVM (**適当なλで**)**

■ 双対問題 $\Longrightarrow$ 不確実性集合アプローチ

ラグランジュ関数：

$$\begin{aligned} &L(\boldsymbol{w}, b, \rho, \boldsymbol{\xi}, \boldsymbol{\alpha}, \mu) \\ &= -2\rho + \frac{1}{m}\sum_{i=1}^{m}\ell(\xi_i) + \sum_{i=1}^{m}\alpha_i(\rho - y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) - \xi_i) + \mu(\|\boldsymbol{w}\|^2 - \lambda^2), \end{aligned}$$

minimax theorem:

$$\inf_{\boldsymbol{w},b,\rho,\boldsymbol{\xi}} \sup_{\boldsymbol{\alpha}\geq\boldsymbol{0},\mu\geq 0} L(\boldsymbol{w}, b, \rho, \boldsymbol{\xi}, \boldsymbol{\alpha}, \mu) = \sup_{\boldsymbol{\alpha}\geq\boldsymbol{0},\mu\geq 0} \inf_{\boldsymbol{w},b,\rho,\boldsymbol{\xi}} L(\boldsymbol{w}, b, \rho, \boldsymbol{\xi}, \boldsymbol{\alpha}, \mu)$$

適当な条件 **(Slater condition** など**)** の下で成立.

$$\sup_{\boldsymbol{\alpha}\geq\mathbf{0},\mu\geq 0}\ \inf_{\boldsymbol{w},b,\rho,\boldsymbol{\xi}} L(\boldsymbol{w},b,\rho,\boldsymbol{\xi},\boldsymbol{\alpha},\mu)$$

$$= \sup_{\boldsymbol{\alpha},\mu\geq 0}\ \inf_{\boldsymbol{w},\boldsymbol{\xi}} \Bigg\{ -\underbrace{\frac{1}{m}\sum_{i=1}^{m}(m\alpha_i\xi_i-\ell(\xi_i))}_{\ell^*(m\alpha_i)\text{がでてくる}} - \sum_{i=1}^{m}\alpha_i y_i \boldsymbol{x}_i^T\boldsymbol{w} + \mu(\|\boldsymbol{w}\|^2-\lambda^2)$$

$$: \sum_{i\in M_p}\alpha_i = \sum_{i\in M_n}\alpha_i = 1,\ \alpha_i \geq 0\Bigg\}$$

$$= -\inf_{\boldsymbol{\alpha}} \Bigg\{ \underbrace{\frac{1}{m}\sum_{i=1}^{m}\ell^*(m\alpha_i)}_{\text{不確実性集合を定義}} + \lambda\Bigg\| \sum_{i:y_i=+1}\alpha_i\boldsymbol{x}_i - \sum_{i:y_i=-1}\alpha_i\boldsymbol{x}_i\Bigg\| :$$

$$\underbrace{\sum_{i:y_i=+1}\alpha_i = \sum_{j:y_j=-1}\alpha_j = 1,\ \alpha_i\geq 0}_{\text{データ点の凸包}}\Bigg\} \qquad \cdots \qquad (\star)$$

損失関数に対応する不確実性集合

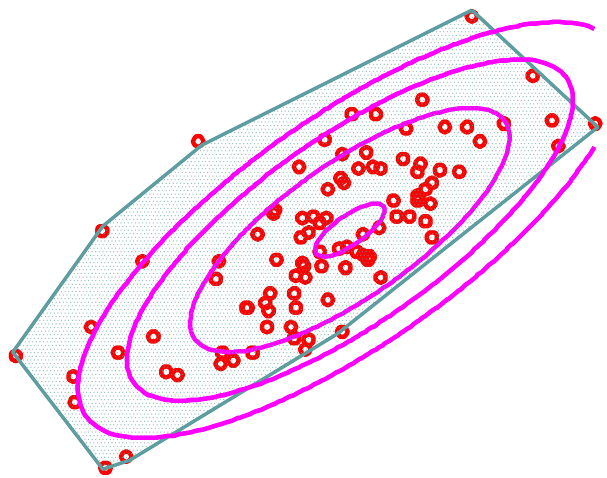
$\ell(z)$：凸・単調非減少・非負

$\Longrightarrow$ **(parametrized)** 不確実性集合 $\mathcal{U}_+[c], \mathcal{U}_-[c],\ c \in \mathbb{R}$.

ℓ^*：ℓ の共役関数 $\ell^*(\alpha) = \sup_{z\in\mathbb{R}}\{\alpha z - \ell(z)\}$.

$$\mathcal{U}_\pm[c] = \left\{ \sum_{i:y_i=\pm 1} \alpha_i \boldsymbol{x}_i \,\middle|\, \alpha_i \geq 0,\ \sum_{i:y_i=\pm 1} \alpha_i = 1,\ \frac{1}{m}\sum_{i:y_i=\pm 1} \ell^*(m\alpha_i) \leq c \right\}$$

データ点の凸包の部分集合



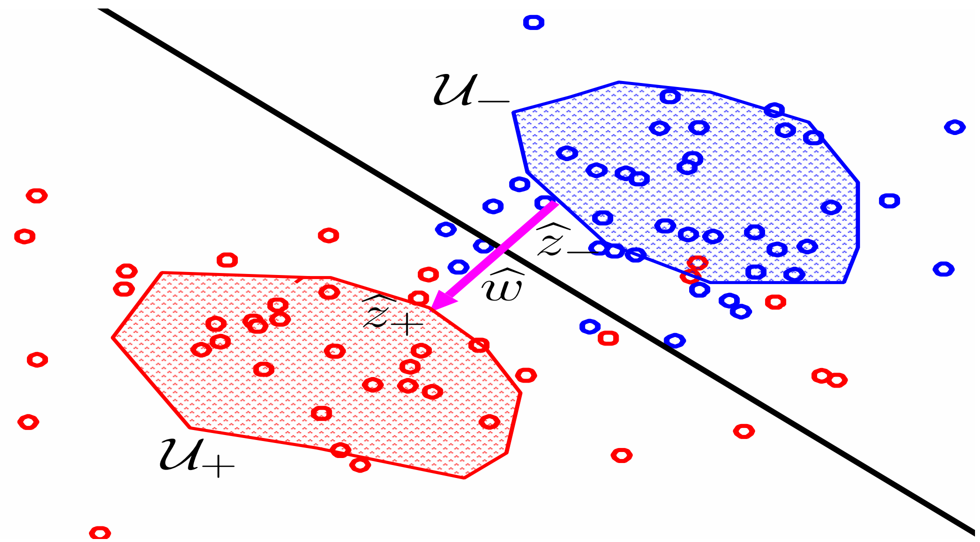
双対問題$(\star)$を不確実性集合で表現

一般化最小距離問題：

$$\min_{\substack{c_+, c_- \\ \boldsymbol{z}_+, \boldsymbol{z}_-}} c_+ + c_- + \lambda \|\boldsymbol{z}_+ - \boldsymbol{z}_-\|, \quad \boldsymbol{z}_+ \in \mathcal{U}_+[c_+],\ \boldsymbol{z}_- \in \mathcal{U}_-[c_-]$$

opt. sol.: $\widehat{\boldsymbol{z}}_+, \widehat{\boldsymbol{z}}_-$. $\quad \widehat{\boldsymbol{w}} = \dfrac{\lambda}{\|\widehat{\boldsymbol{z}}_+ - \widehat{\boldsymbol{z}}_-\|}(\widehat{\boldsymbol{z}}_+ - \widehat{\boldsymbol{z}}_-)$.

$$\Longrightarrow \quad \widehat{f}(\boldsymbol{x}) = \widehat{\boldsymbol{w}}^T \boldsymbol{x} + \widehat{b}$$



不確実性集合アプローチ (修正版)

初期化 ：データから不確実性集合 $\mathcal{U}_{\pm}[c],\ c \in \mathbb{R}$ を定める.

Step 1. 一般化最小距離問題を解く.

$$\min_{c_+, c_-, \boldsymbol{z}_+, \boldsymbol{z}_-} \ c_+ + c_- + \lambda \|\boldsymbol{z}_+ - \boldsymbol{z}_-\|,$$
$$\text{subject to} \ \ \boldsymbol{z}_+ \in \mathcal{U}_+[c_+],\ \ \boldsymbol{z}_- \in \mathcal{U}_-[c_-],\ \ c_\pm \in \mathbb{R}$$

opt. sol.: $\widehat{\boldsymbol{z}}_+,\ \widehat{\boldsymbol{z}}_-$. $\qquad \widehat{\boldsymbol{w}} = \dfrac{\lambda}{\|\widehat{\boldsymbol{z}}_+ - \widehat{\boldsymbol{z}}_-\|}(\widehat{\boldsymbol{z}}_+ - \widehat{\boldsymbol{z}}_-)$

Step 2. $\widehat{b}$ を適当に推定.

判別関数. $\widehat{f}(\boldsymbol{x}) = \widehat{\boldsymbol{w}}^T \boldsymbol{x} + \widehat{b}$

- 既存の方法：パラメータ $c_\pm$ は固定

$$\min_{\boldsymbol{z}_\pm} \|\boldsymbol{z}_+ - \boldsymbol{z}_-\| \quad \text{subject to} \quad \boldsymbol{z}_+ \in \mathcal{U}_+[c_+], \ \ \boldsymbol{z}_- \in \mathcal{U}_-[c_-]$$

- 修正版：**parametrized-**不確実性集合を導入
  - 不確実性集合の大きさ $(c_\pm)$：データへのフィッティングから推定
  - 正則化パラメータ $(\lambda \geq 0)$：**cross validation** などで決める

例：損失関数から定義される不確実性集合

ν-**SVM**：$\ell(z) = \max\{2z/\nu, 0\}, \quad \ell^*(\alpha) = \begin{cases} 0, & \alpha \in [0, 2/\nu], \\ \infty, & \alpha \notin [0, 2/\nu], \end{cases}$

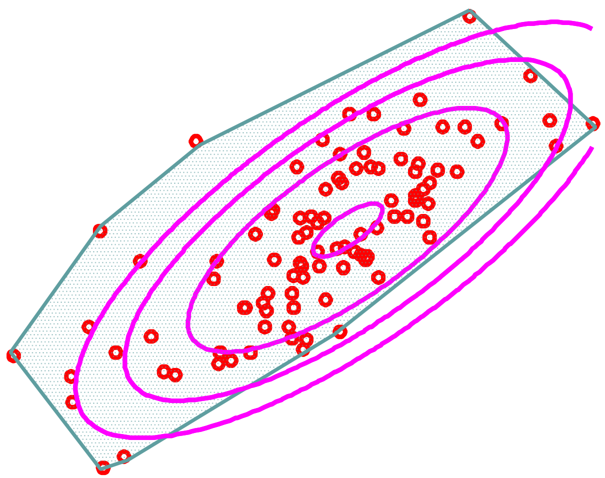
$$\mathcal{U}_\pm[c] = \begin{cases} \left\{ \sum_{i:y_i=\pm 1} \alpha_i \boldsymbol{x}_i \,\middle|\, \sum_{i:y_i=\pm 1} \alpha_i = 1,\ 0 \le \alpha_i \le \dfrac{2}{m\nu} \right\}, & c \ge 0, \\ \emptyset, & c < 0, \end{cases}$$

縮小凸包 ($\mathcal{U}_\pm[c] : c \ge 0$ なら c に依存しない)

$\Longrightarrow$ $c_\pm = 0$ が最適解．$\min_{\boldsymbol{z}_\pm} \|\boldsymbol{z}_+ - \boldsymbol{z}_-\|$ に帰着される．

2乗損失：$\ell(z) = (\max\{1+z, 0\})^2$, $\ell^*(\alpha) = \begin{cases} -\alpha + \dfrac{\alpha^2}{4}, & \alpha \geq 0, \\ \infty, & \alpha < 0. \end{cases}$

$$\mathcal{U}_{\pm}[c] = \left\{ \sum_{i:y_i=\pm 1} \alpha_i \boldsymbol{x}_i \,\middle|\, \sum_{i:y_i=\pm 1} \alpha_i = 1,\ \alpha_i \geq 0,\ \sum_{i:y_i=\pm 1} \alpha_i^2 \leq \frac{4(c+1)}{m} \right\}$$



c が小さいとき：
不確実性集合は楕円形

ベイズリスク一致性

不確実性集合アプローチ(修正版) の統計的性質

- 不確実性集合に対応する損失関数の性質
- 損失に対する解析手法を適用
  - **consistency of regularized kernel classifiers** [10]
  - **classification-calibrated loss** [1]

Theorem 1.

統計モデルは **universal RKHS** $\mathcal{H}$ **(**無限次元統計モデル**)**.

不確実性集合 $\mathcal{U}_{\pm}[c]$ を用いて得られる判別関数

$$\widehat{f}(x) + \widehat{b} \in \mathcal{H} + \mathbb{R}.$$

$\mathcal{U}_{\pm}[c]$ に対応する損失関数が適当な仮定を満たすとき,

$$\Pr\left\{Y \neq \mathrm{sign}(\widehat{f}(X) + \widehat{b})\right\} \xrightarrow{p} \textbf{Bayes error} \ (\text{データ数} \longrightarrow \infty)$$

正則化パラメータ $\lambda = \lambda_m$ は適当なオーダで $\lim_{m \to \infty} \lambda_m = \infty$ とする.

仮定は次項

- 分布に対する仮定： **non-deterministic assumption**.
  「$\forall x,\ \Pr(Y=+1 \mid x)=0 \text{ or } 1$」でない.

- 不確実性集合に対応する損失に適当な仮定：

$$\ell(z): \text{凸,\ 単調非減少,\ 非負値,}$$
$$\underline{\lim_{z\to\infty}\partial\ell(z)=\infty\ \text{など}}$$

  - **truncated-**２乗損失 $(\max\{z,0\})^2$，指数損失 e^z：**O.K.**
  - ヒンジ損失 $\max\{z,0\}$，ロジスティック損失 $\log(1+e^z)$ では **(**いまのところ**)** 証明できていない.

証明の方針

- 不確実性集合アプローチ $\xrightarrow{\text{双対}}$ 損失関数アプローチ.
- 損失の収束性

$$\mathbb{E}_{XY}[\ell(\widehat{\rho} - Y(\widehat{f}(X) + \widehat{b}))] \xrightarrow{p} \inf_{g:\text{可測},\, \rho\in\mathbb{R}} \mathbb{E}_{XY}[\ell(\rho - Yg(X))]$$

  * 再生核ヒルベルト空間 $\mathcal{H}$ 上の経験過程の一様収束性.
  * 普遍カーネル：$\mathcal{H}$ **is dense in** $C(\mathcal{X})$

- 代替損失の理論 **(**を拡張して適用**)**

$$\Pr\{Y \neq \mathrm{sign}(\widehat{f}(X) + \widehat{b})\} \xrightarrow{p} \textbf{Bayes error}$$

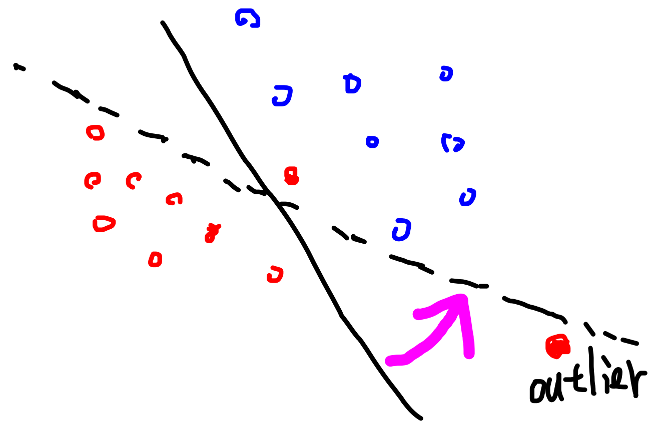
  * ρ は固定：既存の代替損失の理論
  * $\widehat{\rho}$ が確率変数の場合に拡張

まとめ

- 損失関数アプローチと不確実性集合アプローチの関係：
  - 不確実性集合：損失の共役関数のレベル集合
- 不確実性集合を用いる既存の手法を修正
  - 不確実性集合 $\mathcal{U}_{\pm}$ の最小距離問題
  - **Parametrized**-不確実性集合 $\mathcal{U}_{\pm}[c]$ の一般化最小距離問題
- ベイズリスク一致性を証明

その後の展開

- 不確実性集合アプローチ：データ空間での直感的な描像
- 外れ値を含むデータに対する学習アルゴリズムの挙動を解析
  - **classifier can be significantly affected by outliers.**



  - **Outliers should be ignored to prevent overfitting.**
  - 学習アルゴリズムのロバスト化．理論的解析．

References

[1] P. L. Bartlett, M. I. Jordan, and J. D. McAuliffe. Convexity, classification, and risk bounds. Journal of the American Statistical Association, 101:138–156, 2006.

[2] K. P. Bennett and E. J. Bredensteiner. Duality and geometry in SVM classifiers. In Proceedings of International Conference on Machine Learning, pages 57–64, 2000.

[3] D. J. Crisp and C. J. C. Burges. A geometric interpretation of ν-SVM classifiers. In S. A. Solla, T. K. Leen, and K.-R. Müller, editors, Advances in Neural Information Processing Systems 12, pages 244–250. MIT Press, 2000.

[4] T. Kanamori, A. Takeda, T. Suzuki. Conjugate Relation between Loss Functions and Uncertainty Sets in Classification Problems. Journal of Machine Learning Research, vol. 14, pp. 1461-1504, June, 2013.

[5] Gert R.G. Lanckriet, Laurent El Ghaoui, Chiranjib Bhattacharyya, and Michael I.

Jordan. A robust minimax approach to classification. Journal of Machine Learning Research, 3:555–582, 2003.

[6] Michael E. Mavroforakis and Sergios Theodoridis. A geometric approach to support vector machine (svm) classification. IEEE Transactions on Neural Networks, 17(3):671–682, 2006.

[7] J. S. Nath and C. Bhattacharyya. Maximum margin classifiers with specified false positive and false negative error rates. In C. Apte, B. Liu, S. Parthasarathy, and D. Skillicorn, editors, Proceedings of the seventh SIAM International Conference on Data mining, pages 35–46. SIAM, 2007.

[8] B. Schölkopf, A. Smola, R. Williamson, and P. Bartlett. New support vector algorithms. Neural Computation, 12(5):1207–1245, 2000.

[9] I. Steinwart. On the optimal parameter choice for v-support vector machines. IEEE Trans. Pattern Anal. Mach. Intell., 25(10):1274–1284, 2003.

[10] I. Steinwart. Consistency of support vector machines and other regularized

kernel classifiers. IEEE Transactions on Information Theory, 51(1):128–142, 2005.