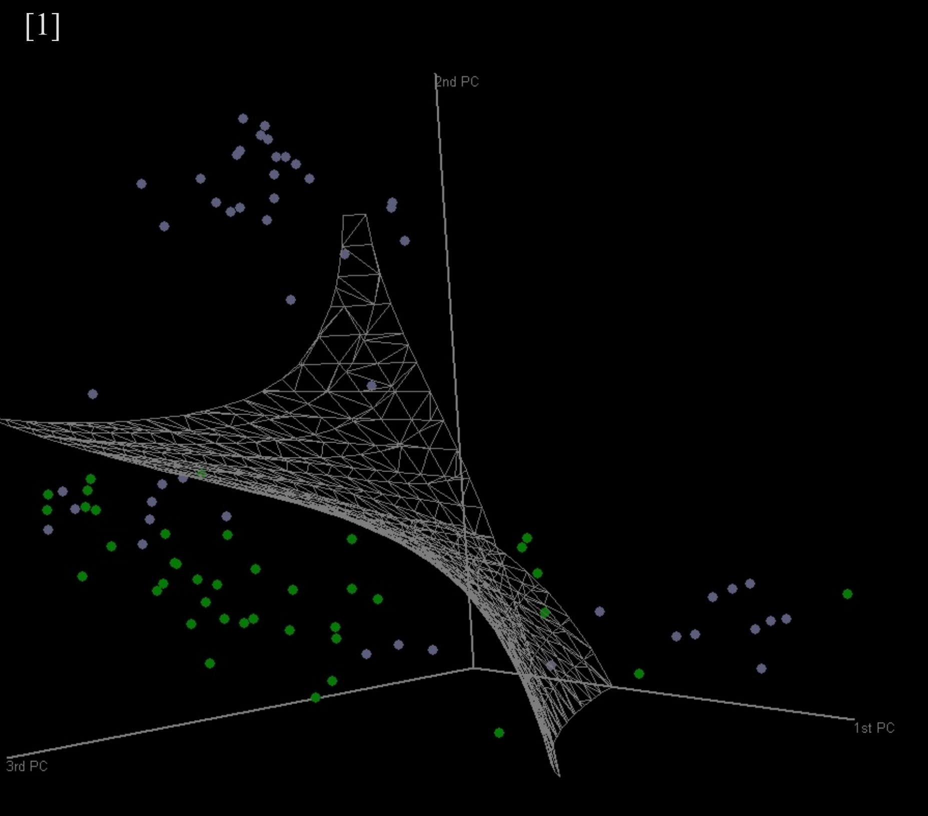
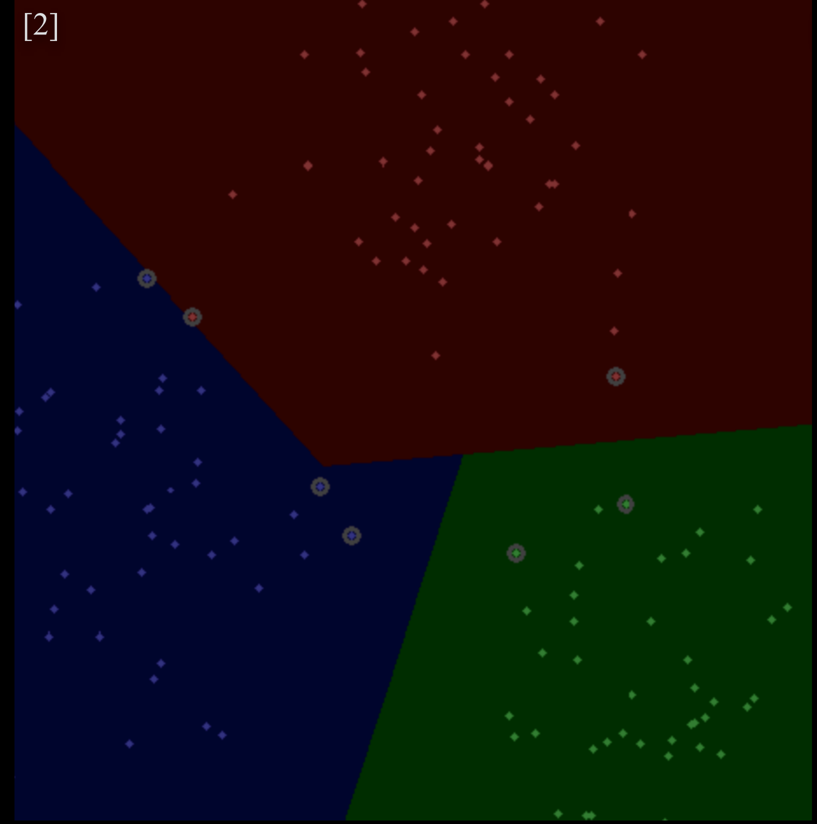
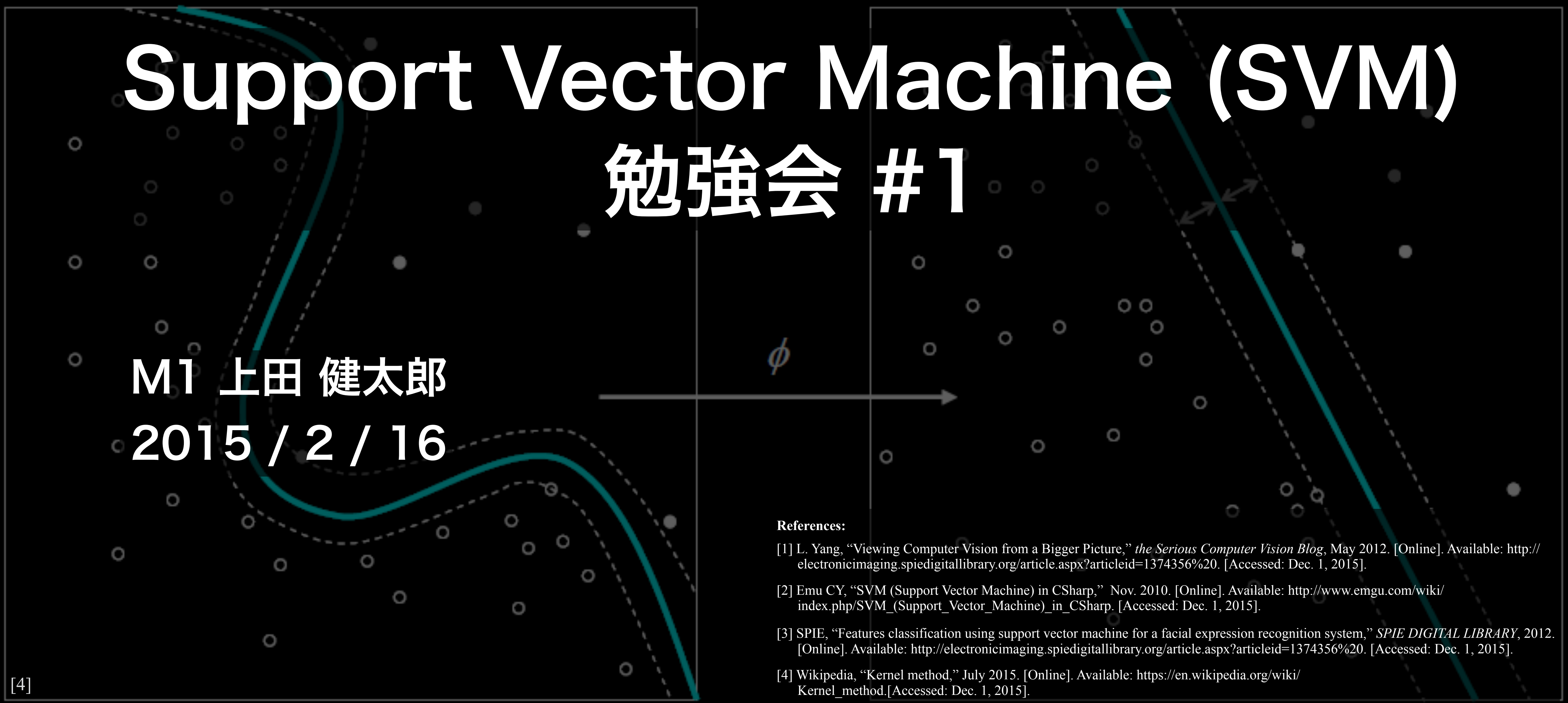






Support Vector Machine (SVM)
勉強会 #1

M1 上田 健太郎

2015 / 2 / 16

References:

[1] L. Yang, "Viewing Computer Vision from a Bigger Picture," *the Serious Computer Vision Blog*, May 2012. [Online]. Available: http://electronicimaging.spiedigitallibrary.org/article.aspx?articleid=1374356%20. [Accessed: Dec. 1, 2015].

[2] Emu CY, "SVM (Support Vector Machine) in CSharp," Nov. 2010. [Online]. Available: http://www.emgu.com/wiki/index.php/SVM_(Support_Vector_Machine)_in_CSharp. [Accessed: Dec. 1, 2015].

[3] SPIE, "Features classification using support vector machine for a facial expression recognition system," *SPIE DIGITAL LIBRARY*, 2012. [Online]. Available: http://electronicimaging.spiedigitallibrary.org/article.aspx?articleid=1374356%20. [Accessed: Dec. 1, 2015].

[4] Wikipedia, "Kernel method," July 2015. [Online]. Available: https://en.wikipedia.org/wiki/Kernel_method.[Accessed: Dec. 1, 2015].

SVMの概要

- 機械学習アルゴリズムの1つ
- SVMは**教師あり**学習
  - 教師あり学習とは… 正解情報がラベリングされたデータを入力とする学習
    - データ例: {(年齢, 腫瘍の大きさ), 0 または 1}

      要素がd個… d次元の学習データ. ベクトルで表す　　教師ラベル (陽性…1，陰性… 0)

    - 他のアルゴリズム… 事例ベース推論，バックプロパゲーション など
  - 教師なし学習とは… 正解情報がラベリングされていないデータを入力とする学習．大量のデータを与え，データ構造を抽出
    - アルゴリズム例… 主成分分析，クラスター分析 など

SVMの歴史

① Vladimirらが線形SVM (後で説明)を発表

- 線形問題しか扱えないため，それほど人気は出ず

② Bernhardらが非線形SVM を発表

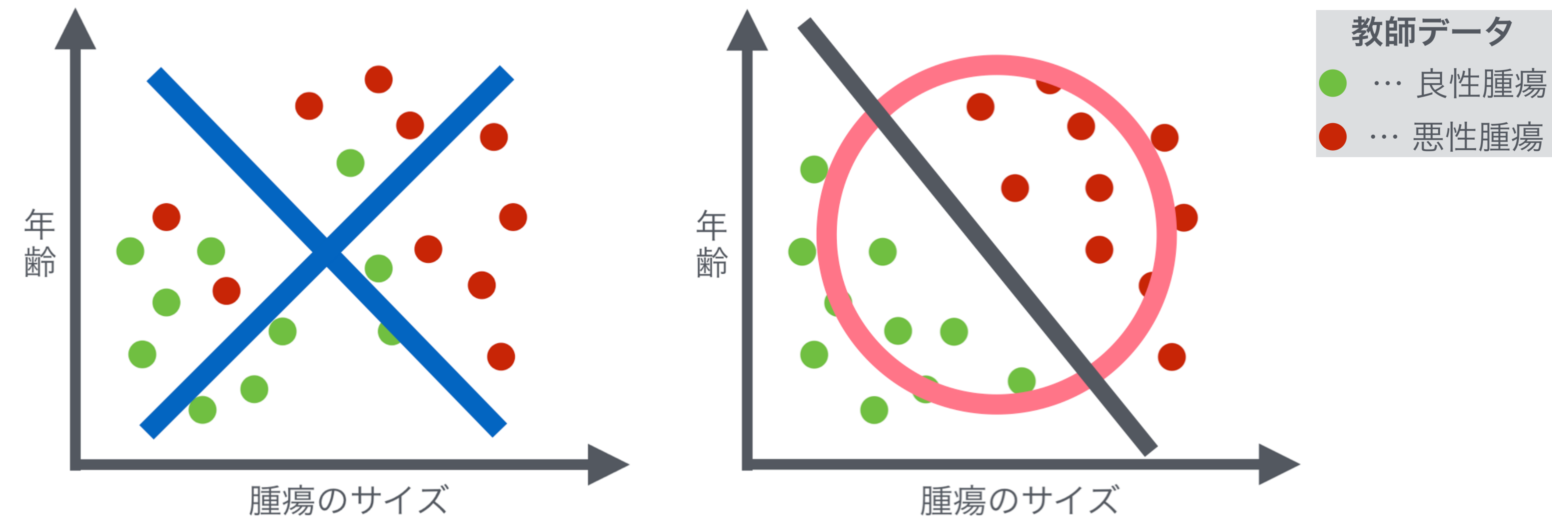
- カーネルトリック という手法を利用
- SVMの適用範囲が広がり，機械学習界隈を賑わした

③ Plattが SMO アルゴリズムを発表

- SVMの計算を効率化するアルゴリズム
- それまでSVMはQP (2時計画法)を利用していたため計算コストが高かった
  - ・しかもQPを解いてくれるソフトウェアは高価
- SVMの実用性が向上し，さらに機械学習界隈を賑わした

SVMの分類

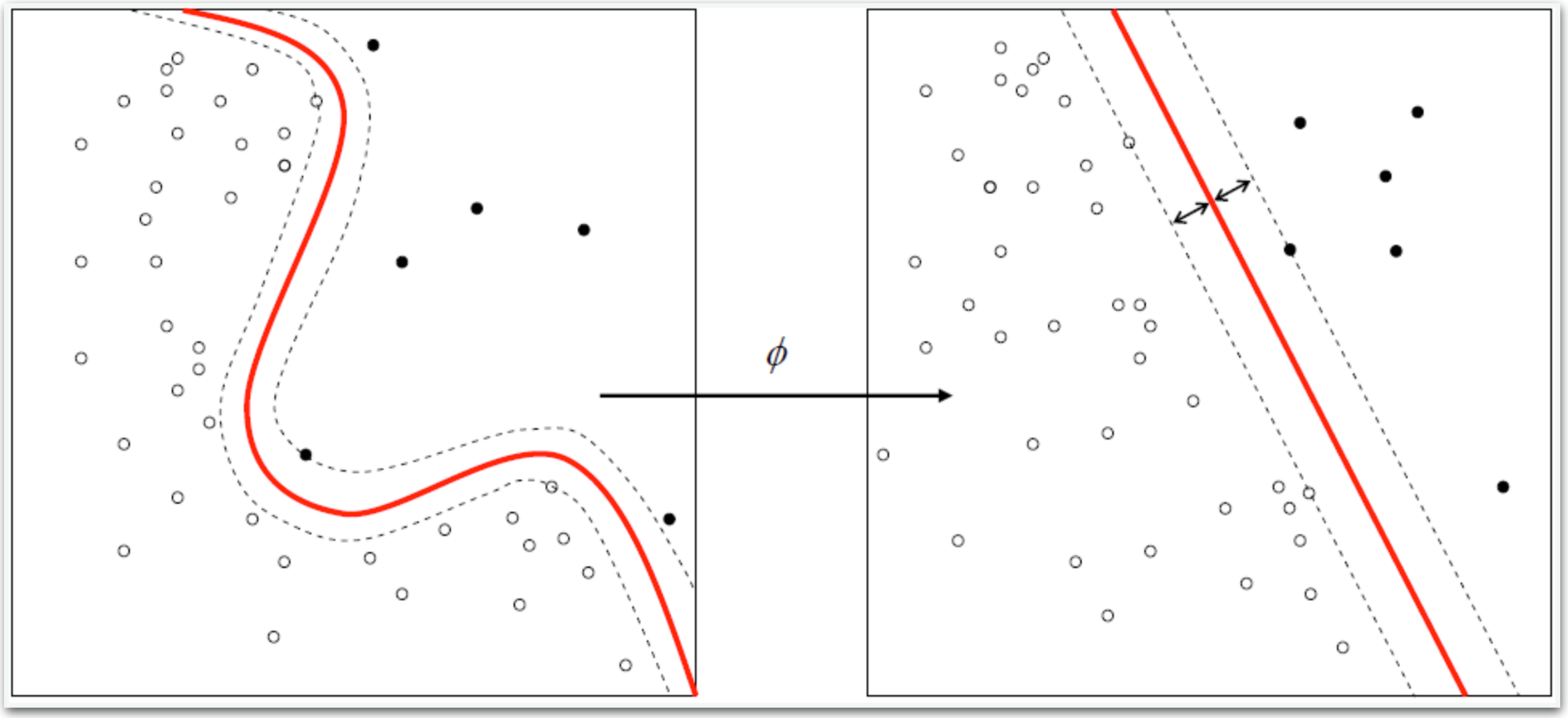
- ハードマージンSVM (今日は主にこれについて話します)
  - 線形分離可能な問題にのみ適用化
  - サンプルプログラムを書きました



- ソフトマージンSVM (第2回 SVM 勉強会で取り上げます)
  - ある程度の線形問題にも適用化
  - 教師データが超平面 (領域の仕切り面) で分離しきれなかった場合ペナルティを与える
  - ペナルティをどの程度重視するかをパラメータ C で表す
  - $C = \infty$ でハードマージンと等価

SVMの分類

- カーネルトリック (第3回 SVM 勉強会で取り上げます)
  - 非線形問題にも適用可能なSVMを実現
  - 線形分離不可能な特徴空間をカーネル関数で線形分離可能な特徴空間に写像し線形分離



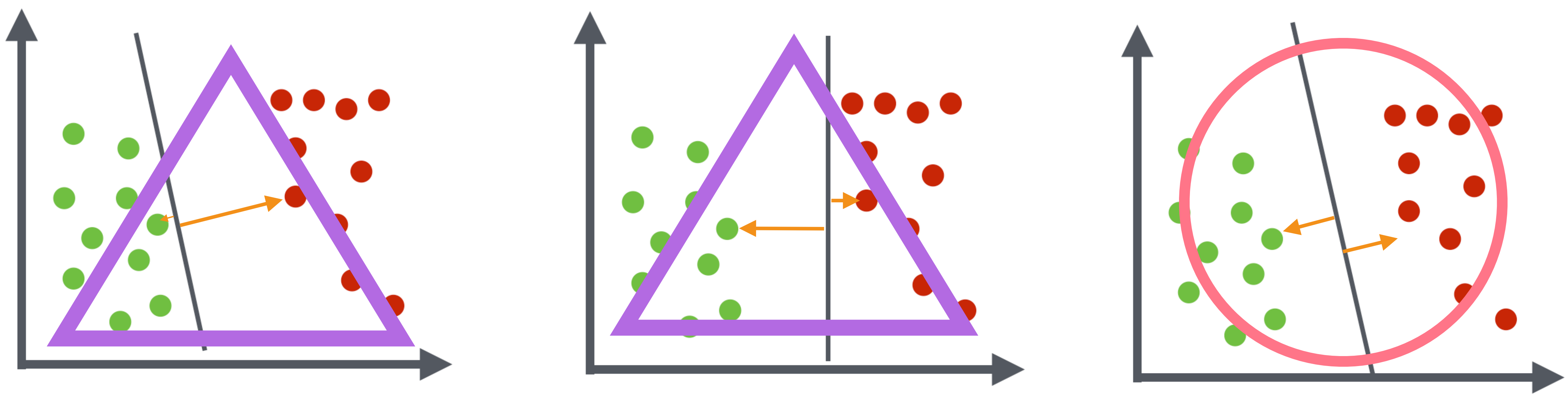
Wikipedia, "Kernel method," July 2015. [Online]. Available: https://en.wikipedia.org/wiki/Kernel_method.[Accessed: Dec. 1, 2015].

・カーネル関数の例… 線形カーネル, ガウスカーネル,
シグモイドカーネル

  - ソフトマージンSVMと組み合わせる場合が多い

ハードマージンSVMとは

- 重要なアイデア… **マージン最大化**
  - ・最近傍の特徴点からの距離を最大化した超平面でデータを線形分離



負例の領域が広く,
汎化能力が低い

負例の領域が狭く,
汎化能力が低い

正例，負例の領域が等しく,
汎化能力が高い

- 2次元なら単純な点と直線の距離(高校数学) でもいい線いきそうだが…
  - ・学習データが多次元だったらどうする?
  - ・高校数学では3次元以上の問題には対応できない

ハードマージンSVM - 定式化①

- 問題を簡単化するために2値分類問題を考える

$\underbrace{\{(x_i, y_i)\}}_{\text{教師付き学習データ}}$　$\underbrace{i = 1, \ldots, N}_{\text{データ数}}$　$\underbrace{y_i = 1,\ \text{または}\ -1}_{\text{正例… 1, 負例... -1}}$

- 識別面は平面なので特徴空間上では以下のように表現可能

$$\omega \cdot x + \omega_0 = 0 \quad (1.1)$$

ここで ω , x は d 次元ベクトル

・なぜ上式が平面を表すか

- ベクトル表現を用いない平面の方程式

$$ax+by+cz+d=0 \quad (1.2)$$

- ベクトル表現を用いた平面の方程式

$$\underbrace{(abc)^T}_{\substack{\text{重み行列}\\ \omega\ \text{とおく}}} \cdot \underbrace{(xyz)}_{x\ \text{とおく}} + \underbrace{d}_{\omega_0\ \text{とおく}} = 0 \quad (1.3)$$

ハードマージンSVM - 定式化②

- i 番目のデータ x_i と識別面との距離 $Dist(x_i)$ は以下のように表せる

$$Dist(x_i) = \frac{|\omega \cdot x + \omega_0|}{\|\omega\|} \quad (1.4)$$

$P_0(x_0, y_0, z_0)$ と $ax+by+cz+d=0$ の距離を導く方程式 $\frac{|ax_0 + by_0 + cz_0 + d|}{\sqrt{a^2 + b^2 + c^2}}$ と同じ

- 識別面に最近傍のデータを代入した際に1となるように重みを調整

$$\min_{i=1,\dots,N} |\omega \cdot x + \omega_0| = 1 \quad (1.5)$$

- (1.4) (1.5) よりデータと識別面の最小距離は以下のようになる

$$\min_{i=1,\dots,N} Dist(x_i) = \min_{i=1,\dots,N} \frac{|\omega \cdot x + \omega_0|}{\|\omega\|} = \frac{1}{\|\omega\|} \quad (1.6)$$

- 以上よりマージンが最大となる識別面を求める問題は以下のとおり

$$\min \frac{1}{2}\|\omega\|^2 \qquad \text{条件: } y_i(\omega \cdot x_i + \omega_0) \geq 1 \quad (1.7)$$

$$i = 1, \dots, N$$

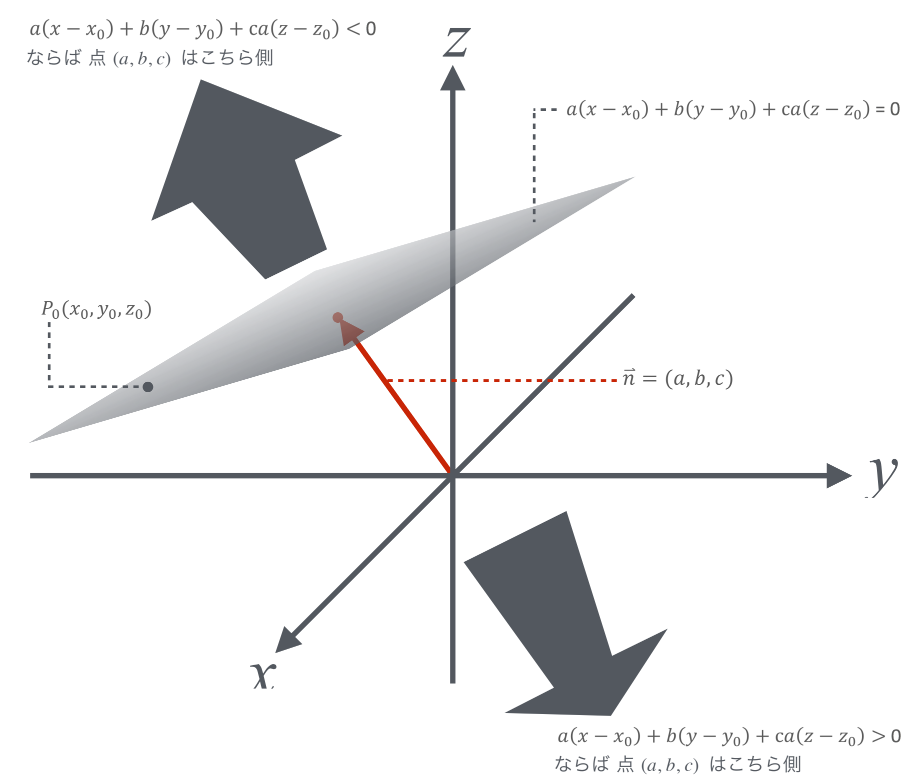
後で微分するので
乗数をかけている

ハードマージンSVM - 定式化②’

- マージンが最大となる識別面を求める問題は以下のとおり (再掲)

$$\min \frac{1}{2}\|\omega\|^2 \quad \text{条件:}\ y_i\left(\omega \cdot x_i + \omega_0\right) \geq 1 \quad i = 1, \ldots, N$$

- なぜ条件式で y_i を掛けているか
  - ・確認: y_i は i 番目の教師データのラベルで，正例で 1，負例で 0 をとる
  - ・高校数学の復習:
    - 点 $P_0(x_0, y_0, z_0)$ を通り法線ベクトル $\vec{n} = (a, b, c)$ に垂直な平面 $a(x - x_0) + b(y - y_0) + ca(z - z_0) = 0$ を考える
    - この時, $a(x - x_0) + b(y - y_0) + ca(z - z_0) > 0$ ならば点 (a, b, c) は**法線ベクトル側の点**
      $a(x - x_0) + b(y - y_0) + ca(z - z_0) < 0$ ならば 点 (a, b, c) は**法線ベクトルの反対側の点**



  - ・条件式は 正例の時は $\underbrace{y_i}_{正}\underbrace{\left(\omega \cdot x_i + \omega_0\right)}_{正} > 0$
    負例の時も $\underbrace{y_i}_{負}\underbrace{\left(\omega \cdot x_i + \omega_0\right)}_{負} > 0$
    となり，1つの式で正例，負例の両方を表せる

ハードマージンSVM - 定式化③

- 制約条件付きの最適化問題にはラグランジュの未定乗数法を利用
- ラグランジュ関数に前ページの式を代入

$$L\left(\omega, \omega_0 \cdot \alpha\right) = -\frac{1}{2}\|\omega\|^2 - \sum_{i=1}^{N} \alpha_i \left(y_i \left(\omega \cdot x_i + \omega_0\right) - 1\right) \quad \text{(1.8)}$$

- KKT条件(不等式制約条件を持つ関数最適化問題で極小値がとらなければならない条件) より

$$\alpha_i \geq 0 \quad \text{(1.9)} \qquad \frac{\partial L}{\partial \omega} = 0 \Rightarrow \omega = \sum_{i=0}^{N} \alpha_i y_i x_i \quad \text{(2.0)} \qquad \frac{\partial L}{\partial \omega_0} = 0 \Rightarrow \sum_{i=0}^{N} \alpha_i y_i = 0 \quad \text{(2.1)}$$

- (1.8) に (1.9) (2.0) (2.1) を代入すると

$$L(\alpha) = \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^T x_j - \sum_{i=1}^{N} \alpha_i \quad \text{(2.2)}$$

- (1.7) の問題を (2.2) を最小化するL を求める問題に置き換えられた
- ω_0 は α_i が 0 でないベクトル(サポートベクトル) より

$$\omega_0 = y_i - \omega \cdot x_i \quad \text{(2.3)}$$

ハードマージンSVM - 定式化③

- 制約条件付きの最適化問題にはラグランジュの未定乗数法を利用
- ラグランジュ関数に前ページの式を代入

$$L(\omega, \omega_0 \cdot \alpha) = -\frac{1}{2}\|\omega\|^2 - \sum_{i=1}^{N} \alpha_i \left(y_i \left(\omega \cdot x_i + \omega_0\right) - 1\right) \quad \text{(1.8)}$$

- KKT条件(不等式制約条件を持つ関数最適化問題で極小値が
  とらなければならない条件) より

α

ラグランジュ未定乗数法により主問題を相対問題に置換する.
後は相対問題を最急降下法や**SMOアルゴリズム**で解く.
P13に参考文献を挙げました

- (1.8) に (1.9) (2.0) (2.1) を代入すると

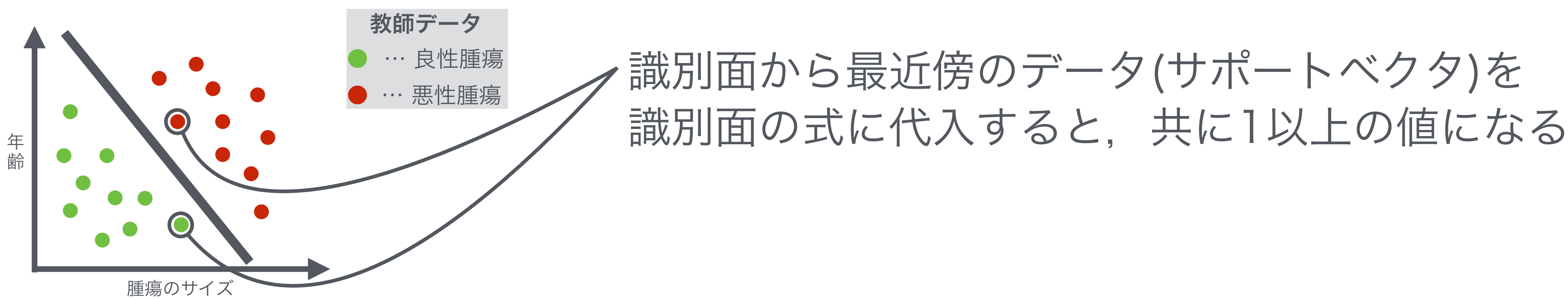
$$L(\alpha) = \frac{1}{2}\sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^T x_j - \sum_{i=1}^{N} \alpha_i \quad \text{(2.2)}$$

- (1.7) の問題を (2.2) を最小化するL を求める問題に置き換えられた
- ω_0 は α_i が 0 でないベクトル(サポートベクトル) より

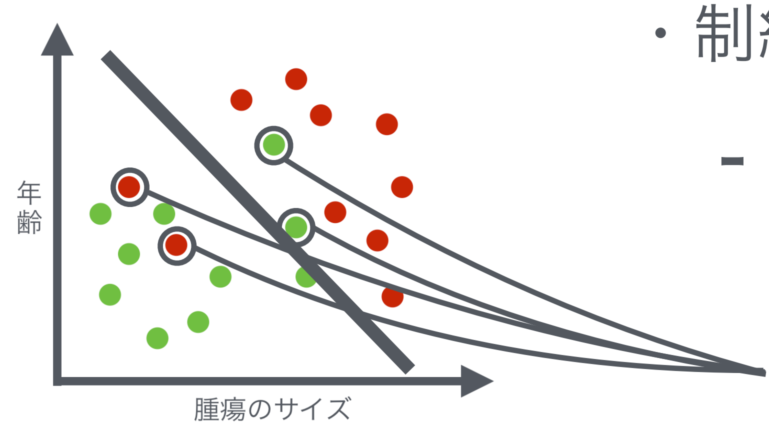
$$\omega_0 = y_i - \omega \cdot x_i \quad \text{(2.3)}$$

ソフトマージンSVMとは

- 重要なアイデア… **線形分離違反の許容度を設定し非線形問題に対応**
  - ・ハードマージンSVMは線形分離可能な問題にしか適用できない
    - 制約条件を $y_i(\omega \cdot x_i + \omega_0) \geq 1$ と定めているため
      $i = 1, \ldots, N$



- 制約条件をどの程度違反して良いかを指定する，スラックと呼ばれる変数を導入



・制約条件の違反の許容度を設定する，とは
- 正例，負例のデータが識別面と逆の方向にどの程度入り込んで良いかを設定する，ということ

ソフトマージンSVM - 定式化①

- ハードマージンSVMの制約条件式に条件違反の許容度を表す
  スラック変数 ς_i を導入

$$y_i\,(\omega \cdot x_i + \omega_0) \geq 1 - \varsigma_i \quad i = 1, \ldots, N \tag{2.1}$$

- このとき，データがマージン(サポートベクトルと識別面の間の区間)
  に入り込んだ場合は　$0 < \varsigma_i \leq 1$

  誤分類した場合は　$\varsigma_i > 1$

- ハードマージンSVMの目的関数にペナルティとしてスラック変数を導入

$$\min \ \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n} \varsigma_i \tag{2.2}$$

  - 定数 C は誤分類されたデータをどの程度重視するかを表す
  - $C = \infty$ でハードマージンSVMと同一

ソフトマージンSVM - 定式化①

- ハードマージンSVMの制約条件式に条件違反の許容度を表すスラック変数 ς_i を導入

$$y_i\,(\omega \cdot x_i + \omega_0) \geq 1 - \varsigma_i \qquad (2.1)$$
$$i = 1, \ldots, N$$

- このとき，データがマージン(サポートベクトルと識別面の間の区間)
に

誤

ハードマージンSVMと同様，
ラグランジュ未定乗数法により主問題を相対問題に置換する.
後は相対問題を最急降下法や**SMOアルゴリズム**で解く.
P13に参考文献を挙げました

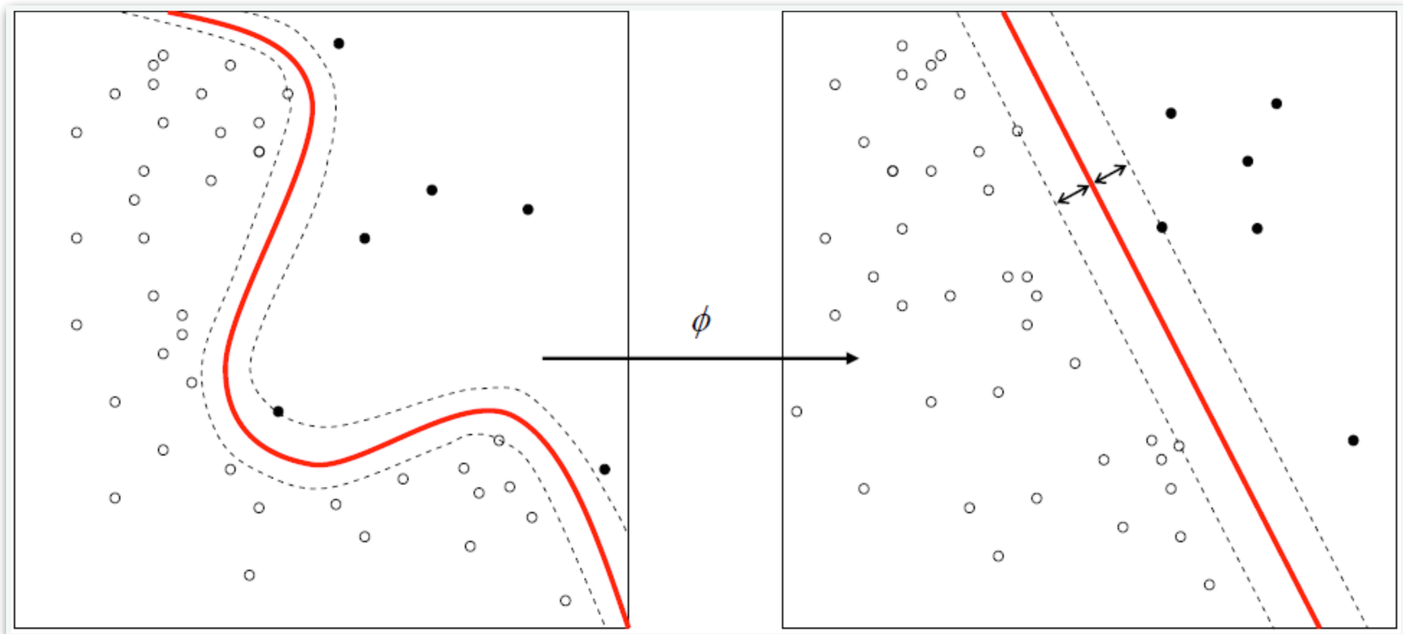
- ハードマージンSVMの目的関数にペナルティとしてスラック変数を導入

$$\min \;\; \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n} \varsigma_i \qquad (2.2)$$

- 定数 C は誤分類されたデータをどの程度重視するかを表す
- $C = \infty$ でハードマージンSVMと同一

非線形SVMとは

- 線形分離不可能な特徴空間をカーネル関数φで線形分離可能な特徴空間に写像し線形分離



Wikipedia, "Kernel method," July 2015. [Online]. Available: https://en.wikipedia.org/wiki/

- 分離後は逆変換してあげればOK
- 目的関数はソフトマージンSVMと同じ

$$\min \ \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}\varsigma_i \quad (2.3)$$

- カーネル関数で写像した特徴空間を利用するため制約条件は異なる

$$y_i(\omega^t\varphi(x_i) + \omega_0) - (1 - \varsigma_i) \geq 0 \quad (2.4)$$

非線形SVMとは

・線形分離不可能な特徴空間をカーネル関数φで線形分離可能な特徴空間に写像し線形分離

ハードマージンSVM，ソフトマージンSVMと同様に，
ラグランジュ未定乗数法により主問題を相対問題に置換する.
後は相対問題を最急降下法や**SMOアルゴリズム**で解く.
P13に参考文献を挙げました

・分

・目的関数はソフトマージンSVMと同じ

$$\min \; \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n} \varsigma_i \quad (2.3)$$

・カーネル関数で写像した特徴空間を利用するため制約条件は異なる

$$y_i(\omega^t \varphi(x_i) + \omega_0) - (1 - \varsigma_i) \geq 0 \quad (2.4)$$

ぜひ参考文献を読んでください!

- SVMのアルゴリズム解説

  http://shogo82148.github.io/homepage/memo/algorithm/svm/
  (サンプルプログラムもあります)

  http://sudillap.hatenablog.com/entry/2013/04/08/235610

  http://www.slideshare.net/thorikawa/2-34svm

  http://www.neuro.sfc.keio.ac.jp/~masato/study/SVM/index.htm

  http://www.slideshare.net/sleepy_yoshi/svm-13435949

  http://www.slideshare.net/ShinyaShimizu/ss-11623505

  https://github.com/levelfour/machine-learning-2014/wiki/第4回---非線形写像とカーネル関数

- 最急降下法，SMOアルゴリズム解説

  http://shogo82148.github.io/homepage/memo/algorithm/steepest-descent/

  http://convexbrain.osdn.jp/cgi-bin/wifky.pl?p=SMO%A5%A2%A5%EB%A5%B4%A5%EA%A5%BA%A5%E0

  http://sssslide.com/www.slideshare.net/sleepy_yoshi/smo-svm

  http://shogo82148.github.io/homepage/memo/algorithm/svm/svm-detail.html

- ラグランジュ未定乗数法

  http://eman-physics.net/analytic/lag_method.html

次回(18日木曜日，最終回)予告

- LivSvm の導入
  - Iris 分類くらいはやってみたい
  - LibSvm利用方法，データ前処理の話題が中心になりそうです
- カーネルSVM独力実装
  - ソフトマージンSVMから容易に拡張可能説…
  - 1日粘ってダメだったら LibSvm に注力します…
- GitHubにサンプルプログラム&資料をアップロードします
  - https://github.com/KentaroUeda