数理計画法とサポートベクターマシン

矢島 安敏 (yasutosi@me.titech.ac.jp)
東京工業大学 経営工学専攻
〒 152-8552 目黒区大岡山 2-12-1

1 はじめに

近年 Support Vector Machine(SVM) を用いた判別法が, 文書の分類 [6, 9] や画像の認識 [4, 14, 13] といったさまざまの分野に応用され, 非常に有力な判別法と考えられるようになってきている. SVM は従来の判別法と比べると, 最適化問題を解く必要があるなど計算が複雑で, 特に大規模データに対しては実用的な意味で適用できないと考えられていた時期もあった. しかし, 計算機能力の進歩と様々の最適化技術を取り入れた高速アルゴリズムの登場で, データマイニング手法の一つとして, 最近ではマイニングパッケージにも取り入れられるようになってきている.

本稿では, まず次節において, Vapnik による標準的な SVM の定式化を示すとともに, 現在まで提案されている幾つかのバリエーションを紹介する. これらに共通して用いられているアイディアは, ridge あるいは lasso と呼ばれる回帰分析の分野では古くから用いられているものであることを紹介する. 第 3 節では, 特に線形判別関数を構成する場合に注目し, 大規模データにも適応可能なアルゴリズムについて述べる. 4 節では, これらのアルゴリズムを非線形な判別にも適用する方法について考える.

2 SVMによる判別

N 次元空間 $\mathbb{R}^N$ に M 個のデータ (点) が与えられているとする. 各点 $j = 1, 2, \ldots, M$ には 2 値のラベル $y_j \in \{-1, +1\}$ が与えられている. このとき, ラベルの値にしたがって点を判別する 2 クラスの判別問題を考える. M 個の点は M 行 N 列の行列 A により表されているとする. すなわち各データ j を A の第 j 行ベクトル A_j と表すことにする. また, 各ラベルの値を要素とする M 次元ベクトルを $y = (y_1, y_2, \ldots, y_M)^T$, これを対角成分とする M 次対角行列を Y と定義する.

SVM では線形関数を用いた判別を行う. N 次元の法線ベクトル w および実数 b で定まる線形関数を $f(x) = x^T w - b$ とすれば, SVM の目的は与えられたデータ A およびラベル y にしたがって,

$$f(A_j) = A_j w - b \begin{cases} > 0 & \text{if } y_j = 1, \\ < 0 & \text{if } y_j = -1, \end{cases} \qquad j = 1, 2, \ldots, M$$

となる $f(x)$ を導出することである.

一般的には, 与えられた点が線形関数で完全に分離できるとは限らないので, 非負のスラック変数 $\xi_j \geq 0$, $j = 1, 2, \ldots, M$ を導入し, 以下の制約条件

$$\begin{array}{ll} A_j w - b + \xi_j \geq 1 & \text{if } y_j = 1, \\ A_j w - b - \xi_j \leq -1 & \text{if } y_j = -1 \end{array} \tag{2.1}$$

のもと, スラック変数の総和 $\sum_{j=1}^{M} \xi_j$ が最小となる線形関数を考える. 一方で, SRM 原理 [17] によれば, $\|w\|$ の大きな関数を構成すると, 汎か能力の低い判別となってしまう恐れがあるとされている. そこで, Vapnik[17] は, $\sum_{j=1}^{M} \xi_j$ と $\|w\|$ 双方を最小化するために, 次の二次計画問題を解きベクトル $w \in \mathbb{R}^N$ および b を求めるこ

とを提案した.

$$(2.2)\quad \left|\begin{array}{ll} \text{最小化} & \|w\|_2^2 + C\sum_{j=1}^{M}\xi_j \\ \text{制約} & Y(Aw-be)+\xi \geq e,\ \ \xi \geq 0, \end{array}\right.$$

ここで, $\xi = (\xi_1, \xi_2, \ldots, \xi_M)^T \in \mathbb{R}^M$, e は要素がすべて 1 のベクトル, また C はあらかじめ設定された正の定数で, $\|w\|_2^2$ と $\sum_{j=1}^{M}\xi_j$ とのバランスをコントロールするパラメータである.

通常は, この問題の双対問題 [5, 17] を考え最適化を行う. 行列 K を M 次の対称行列で $K = AA^T$ となるもの, $\alpha \in \mathbb{R}^M$ を双対変数とすれば, 問題 (2.2) の双対問題は

$$(2.3)\quad \left|\begin{array}{ll} \text{最大化} & -\frac{1}{2}\alpha^T YKY\alpha + e^T\alpha \\ \text{制約} & y^T\alpha = 0, \quad 0 \leq \alpha \leq Ce \end{array}\right.$$

と書くことができる.

今, 式 (2.1) を

$$(2.4)\quad \xi_j = \min\{\ 0,\ 1-y_j(A_jw-b)\ \} = \min\{\ 0,\ y_j(y_j - f(A_j))\ \}$$

と考えれば, ξ_j は, データ A_j が誤って判別された場合に与えられるペナルティであり,

$$(2.5)\quad f(x) = y_j$$

となる関係 ($f(x) = 0$ ではない) からの差に相当していると見なすことができる.

ペナルティの与え方によって, 次のようなバリエーションが提案されている. 例えば, $\|w\|_2^2$ とスラック変数の 2 乗の総和 $\sum_{j=1}^{M}\xi_j^2$ を考えるならば, 次のような最小化問題が得られる.

$$(2.6)\quad \left|\begin{array}{ll} \text{最小化} & \|w\|_2^2 + C\sum_{j=1}^{M}\xi_j^2 \\ \text{制約} & Y(Aw-be)+\xi \geq e. \end{array}\right.$$

このように定式化をすれば ξ_j の非負制約は不要となる. ゆえに, この双対問題は

$$(2.7)\quad \left|\begin{array}{ll} \text{最大化} & -\frac{1}{2}\alpha^T Y(K+\frac{1}{C}I)Y\alpha + e^T\alpha \\ \text{制約} & y^T\alpha = 0, \quad 0 \leq \alpha \end{array}\right.$$

となる. ただし I は単位行列である. 双対問題 (2.3) と比較すると, 目的関数の二次項が $K + \frac{1}{C}I$ となり, また, ξ の非負制約が無いことで, α の上限の制約がなくなっている.

さらに, 式 (2.5) を満たすよう f を定めるのであれば, 単純に差の二乗和の最小化が考えられ,

$$(2.8)\quad \left|\begin{array}{ll} \text{最小化} & \|w\|_2^2 + C\sum_{j=1}^{M}\xi_j^2 \\ \text{制約} & Y(Aw-be)+\xi = e \end{array}\right.$$

を得る. 等式制約より ξ が消去できるので, この問題は, 単に二次関数

$$\|w\|_2^2 + C\|y - (Aw-be)\|_2^2$$

の最小化となる. これは ridge 回帰 [8] と呼ばれるもので, 回帰の予測能力や結果の説明力を高めるために古くから用いられている手法である. 最初に示した SVM の定式化 (2.2) は, ridge 回帰のペナルティ項 (ξ) を二乗誤

差から式 (2.4) で与えられるような 1 次関数に置き換えたものと見なすことができる. また, $\|w\|_2^2$ を 1 ノルム $\|w\|_1$ に置き換えた

$$\|w\|_1 + C\|y - (Aw - be)\|_2^2$$

は lasso (Least Absolute Shrinkage and Selection Operator) [15] と呼ばれ, やはり汎か能力向上の効果があるとされている. 先ほど同様に, ペナルティ項を式 (2.4) のタイプの線形関数に置き換えるのなら,

$$(2.9)\quad \left| \begin{array}{ll} \text{最小化} & \|w\|_1 + C\sum_{j=1}^{M} \xi_j \\ \text{制約} & Y(Aw - be) + \xi \geq e, \quad \xi \geq 0, \end{array} \right.$$

と線形計画問題が得られる.

3 SVMに対する最適化手法

問題 (2.3) などはいずれも凸二次計画問題であり, 一般的には内点法 [16] などにより効率的に最適化が可能であると考えられる. しかし, データマイニングに見られる問題では, データの次元 (N) はさほど大きくないものの, データの数 (M) が極めて大きな場合を扱わなくてはならない. 例えば双対問題 (2.3) では, $M \times M$ の大型で稠密な行列 K を扱わなくてはならない. M が数万を超えれば, K を主記憶に配列で保持することはできなくなってしまう. また, 問題 (2.3) の最適解では多くの変数が 0 であることが期待されるので, このような特殊構造を用いることが効率的解法の鍵となると考えられ, 幾つかの手法が提案されている [10, 12, 3]. これらのアルゴリズムは, 最適解での 0 の要素の減少に伴い効率が低下すること, また, 高い精度で最適解を得ることが困難であるなど問題点もあるが, 実用的観点からは効率的な手法であると考えられる.

現実の問題の中には, M は大きいものの, それに比べて N が小さな問題も存在する. このような場合には双対問題 (2.3) を解く事が必ずしも得策ではない. ここでは, データの次元 N がデータ数 M に比べ非常に小さい場合有効であると考えられる Lagrangian Support Vector Machine (LSVM) [11] について簡単に述べる.

まず, 定式化 (2.6) をさらに変形させ, 目的関数に b^2 を加えた

$$(3.10)\quad \left| \begin{array}{ll} \text{最小化} & \|w\|_2^2 + b^2 + C\sum_{j=1}^{M} \xi_j^2 \\ \text{制約} & Y(Aw - be) + \xi \geq e. \end{array} \right.$$

を考える. この場合, 目的関数は strictly な凸となり, また双対問題は,

$$(3.11)\quad \left| \begin{array}{ll} \text{最大化} & -\frac{1}{2}\alpha^T Y(AA^T + ee^T + \frac{I}{C})Y\alpha + e^T\alpha \\ \text{制約} & 0 \leq \alpha, \end{array} \right.$$

と凸二次関数を非負象限で最適化する問題となる. 以降簡単のため, M 行 $N+1$ 列の行列を $H = Y[A - e]$, また $Q = HH^T + \frac{I}{C}$ と置き, 問題 (3.11) を最小化問題として

$$\text{最小化}\left\{ \frac{1}{2}\alpha^T Q\alpha - e^T\alpha \mid \alpha \geq 0 \right\}$$

と書くことにする. このとき KKT 条件から, 次の反復式

$$(3.12)\quad \alpha^{k+1} = Q^{-1}\left(e + \max\{0, Q\alpha^k - e - \alpha^k\mu\}\right)$$

を導くことができ, 実数 μ を適当に定めれば, 点列 $\{\alpha^k\}$ は最適解 α^* に収束することが示されている [11].

ここでポイントとなる点は, 反復を行うためには $Q = HH^T + \frac{I}{C}$ の逆行列を 1 回計算すればよいこと, さらに,

$$(3.13)\quad Q^{-1} = \left(HH^T + \frac{I}{C}\right)^{-1} = C\left(I - H\left(H^TH + \frac{I}{C}\right)^{-1}H^T\right)$$

となる関係を使えば, $H^TH + \frac{I}{C}$ の逆行列, すなわち N 次行列の逆行列を計算すればよく, データ数 M が大きくても Q^{-1} を算出することが可能である. また, 反復計算 (3.12) の実行には行列 H を主記憶に保持すればよく, H の疎性 (もしあれば) も利用可能である.

さらに, N が小さい状況ならば, たとえ M が相当に大きな場合でも, 線形計画問題 (2.9) であれば, 標準的な単体法や内点法で解くことも可能である. また, $K = AA^T$ となることを使えば, 二次計画問題 (2.3) を内点法で解くことは, 線形計画問題 (2.9) を解く場合と計算上大きな差はなく, 大規模問題にも十分適応可能である [7].

4 カーネルを使った非線形な判別関数の構成

SVM が多くの問題に対して高い判別力を示すことができるのは, 双対問題 (2.3) などを用いて非線形な判別関数を構成できる点にある. 非線形な判別関数を構成するためには, まず, 適当な非線形写像 $\phi : \mathbb{R}^N \to \mathcal{F}$ を使い各データ A_j をより高い次元の特徴空間 $\mathcal{F}$ へと射影する. 射影された $\mathcal{F}$ の元 $\phi(A_1), \phi(A_2), \ldots, \phi(A_M)$ に対して最適化問題 (2.3) を解けば, $\mathcal{F}$ 上での線形な判別関数が得られる. これをもとの空間で見れば結果として非線形な判別が行われることになる. ここで双対問題 (2.3) を見ると, $\mathcal{F}$ 上の内積からなるグラム行列 $\mathcal{K} \in \mathbb{R}^{M\times M}$ が必要でり, 特徴空間の元 $\phi(A_j)$ は定式化の中に陽には現れてこない.

そこで, SVM ではカーネル関数と呼ばれる特殊な関数を用い $x, x' \in \mathbb{R}^N$ から直接, $\mathcal{F}$ の元 $\phi(x), \phi(x)$ の内積 $\langle \phi(x), \phi(x') \rangle$ を求め, 双対問題の最適化により非線形の判別を実現する. よく用いられる代表的なカーネル関数として, polynomial kernel $\mathcal{K}(x, x') = \left(x^Tx' + 1\right)^d$, RBF kernel $\mathcal{K}(x, x') = \exp\left(-\|x - x'\|^2/\sigma^2\right)$, あるいは sigmoid kernel $\mathcal{K}(x, x') = \tanh\left(\kappa x^Tx' - \Theta\right)$ (ただし d は自然数のパラメータ, σ, κ, Θ は実数のパラメータである), などがあり, さまざまの分野 [4, 14, 9] で用いられ, 有効であると考えられている.

このように, SVM では非線形の判別関数を構成するためには, 双対問題 (2.3) や (2.7) のように定式化される二次計画問題の最適化が必要で, しがって稠密で大規模なグラム行列 $\mathcal{K}$ を持った最適化問題を解かなくてはならない. また, LSVM の場合でも式 (3.13) を使うことができず, Q^{-1} を陽に保持しなくてはならない. そこで, $\phi(A_j)$ を低い次元の実ベクトルとして近似的に表現し, 線形判別のアルゴリズムを用いることを考える.

そのために, まずグラム行列 $\mathcal{K}$ と非線形写像 ϕ との関係について考える. 実数 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M \geq 0$ を $\mathcal{K}$ の固有値, また $p_1, p_2, \ldots, p_M \in \mathbb{R}^M$ を対応した固有ベクトルで長さ 1 に正規化されているものとする. このとき, 0 でない固有値の中から, 大きなもの $S(< M)$ 個とそれに対応する固有ベクトルを使い, M 行 S 列の行列

$$D_S = \left[\sqrt{\lambda_1}p_1\sqrt{\lambda_2}p_2\cdots\sqrt{\lambda_S}p_S\right]$$

を定める. 行列 $D_SD_S^T$ はランクが S の行列で Frobenius norm の意味で $\mathcal{K}$ の最良の近似となっている.

D_S の各要素と空間 $\mathcal{F}$ の間には次のような関係を導くことができる. 行列 D_S の $j-k$ 要素を d_{jk} とし, 以下の様にして定まる $\mathcal{F}$ の元

$$(4.14)\quad \mathcal{V}_k = \frac{\sum_{j=1}^M d_{jk}\phi(A_j)}{\lambda_k},\quad k = 1, 2, \ldots, S$$

を考える. このとき次の補題が成立する.

補題 4.1 特徴空間 $\mathcal{F}$ の内積を $\langle \cdot, \cdot \rangle$ と書くことにする．$\mathcal{F}$ の元 $\{\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_S\}$ は正規直交基底となる．

証明 行列 D_S の第 k 列ベクトルを $d_k \in \mathbb{R}^M$ とする．このとき，簡単な計算により，任意の $k, k' = 1, 2, \ldots, S$ に対して

$$\langle \mathcal{V}_k \, , \, \mathcal{V}_{k'} \rangle \;=\; \frac{1}{\lambda_k \lambda_{k'}} \sum_{j=1}^{M} \sum_{j'=1}^{M} d_{jk} d_{j'k'} \langle \phi(A_j) \, , \, \phi(A_{j'}^T) \rangle = \frac{1}{\lambda_k \lambda_{k'}} d_k^T \mathcal{K} d_{k'}$$

と変形できる．明らかに，もし $k \neq k'$ ならば $d_{k'}^T \mathcal{K} d_k = 0$ となるので，$\langle \mathcal{V}_k \, , \, \mathcal{V}_{k'} \rangle = 0$ であり，$d_k^T \mathcal{K} d_k = \lambda_k \|d_k\|^2 = \lambda_k^2$ より $\langle \mathcal{V}_k \, , \, \mathcal{V}_k \rangle = 1$ を得る． □

そこで，基底

$$\mathcal{V} = \{\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_S\} \tag{4.15}$$

で張られる $\mathcal{F}$ の部分空間を $\mathcal{F}_S$ と表す．このとき，任意の A_j，$j = 1, 2, \ldots, M$ に対してベクトル $\phi(A_j)$ の $\mathcal{F}_S$ への射影は

$$\langle \phi(A_j) \, , \, \mathcal{V}_1 \rangle \mathcal{V}_1 + \langle \phi(A_j) \, , \, \mathcal{V}_2 \rangle \mathcal{V}_2 + \cdots + \langle \phi(A_j) \, , \, \mathcal{V}_S \rangle \mathcal{V}_S \tag{4.16}$$

と表現することができる．また，ベクトル (4.16) の基底 $\mathcal{V}$(4.15) に関する座標ベクトル

$$(\langle \phi(A_j) \, , \, \mathcal{V}_1 \rangle \; \langle \phi(A_j) \, , \, \mathcal{V}_2 \rangle \; \cdots \; \langle \phi(A_j) \, , \, \mathcal{V}_S \rangle)$$

は D_S の第 j 行ベクトルに他ならない．

さらに一般的に，任意の点 $x \in \mathbb{R}^N$ に対し，非線形写像 $\phi(x)$ の基底 $\mathcal{V}$ に関する S 次元の座標ベクトルを $[x]_{\mathcal{V}}$，すなわち

$$[x]_{\mathcal{V}} = (\langle \phi(x) \, , \, \mathcal{V}_1 \rangle \; \langle \phi(x) \, , \, \mathcal{V}_2 \rangle \; \cdots \; \langle \phi(x) \, , \, \mathcal{V}_S \rangle)^T \in \mathbb{R}^S$$

と記すことにする．$[x]_{\mathcal{V}}$ の第 k 成分は

$$\langle \phi(x) \, , \, \mathcal{V}_k \rangle = \left\langle \phi(x) \, , \, \frac{\sum_{j=1}^{M} d_{jk} \phi(A_j)}{\lambda_k} \right\rangle = \frac{\sum_{j=1}^{M} d_{jk} \mathcal{K}(x \, , \, A_j)}{\lambda_k} \tag{4.17}$$

と関数 $\phi(x)$ を陽に定めることなく，カーネル関数 $\mathcal{K}(x \, , \, A_j)$ のみより求めることが可能である．

以上のように，カーネルから作られたグラム行列 $\mathcal{K}$ の固有値と固有ベクトルを使えば，$\mathcal{F}$ の元 $\phi(A_j)$ を S 次元の実ベクトル $[A_j]_{\mathcal{V}}$ として近似的に表現することが可能なり，この S 次元空間で線形判別を求めることで，結果的にもとの N 次元空間での非線形判別を行うことが可能となった．筆者等は，データとして A の代わりに D_S を用い S 次元空間での線形判別を線形計画問題 (2.9) を解き求めところ，通常の SVM による非線形判別に近い判別力のある関数を，効率よく構成できることを確認している [19]．また，LSVM による定式化を使うのであれば，式 (3.13) の最終項の逆行列部分が

$$\left(H^T H + \frac{I}{C}\right)^{-1} = \left(\begin{bmatrix} D_S^T \\ -e^T \end{bmatrix} [D_S - e] + \frac{I}{C}\right)^{-1} = \left(\begin{bmatrix} D_S^T D_S & -D_S^T e \\ -e^T D_S & M \end{bmatrix} + \frac{I}{C}\right)^{-1}$$

となるが，固有ベクトルの性質より $D_S^T D_S = I$ であることを使えば，逆行列の計算も容易に行うことができる．

5 おわりに

本稿では, SVMのいくつバリエーションをその定式化とともに紹介した. Vapnikの提案したSVMでは, カーネルを用いた非線形判別を行うため, 二次の双対問題 (2.3) が導入された. しかし, 線形な判別を行うのであれば, 必ずしもこの二次計画問題を解く必要はなく, より単純な問題 (3.11), あるいは線形計画問題 (2.9) でも十分に能力高い判別関数を構成することが可能である. さらに, 上で説明したように特徴空間の元を低い次元に近似的に表現することを行えば, 非線形な判別関数もLSVMや線形計画法の最適化によって構成可能である.

本稿では, 2クラスの判別問題のみを取り上げたが, 3クラス以上の多クラスの分類問題に対しても区分的に線形な判別関数 [1] を定めることで, 2クラスの場合とほぼ同様な定式化が可能である. 双対問題が凸二次計画問題 [2] として定式化されることを使い, 非線形な多クラス判別を行うM-SVMという手法も提案されている. しかし, M-SVMで使われる二次計画問題は (2.3) より複雑で, 問題 (2.3) の場合のような特殊なアルゴリズムの構築は望めない. そこで, 特徴空間 $\mathcal{F}$ の元を S 次元の点で近似的に表現し線形判別を行えば, 例えば, 線形計画問題を解くことで非線形な多クラスの判別が可能となる [18].

参考文献

[1] K. P. Bennett and O. L. Mangasarian, *Multicategory discrimination via linear programming*, Optimization Methods and Software, 3 (1993), pp. 27–39.

[2] E. J. Bredensteiner and K. P. Bennett, *Multicategory classification by support vector machines*, Computational Optimization and Applications, 12 (1999), pp. 53–79.

[3] R. Collobert and S. Bengio, *SVMTorch: Support vector machines for large-scale regression problems*, Journal of Machine Learning Research, 1 (2001), pp. 143–160.

[4] C. Cortes and V. Vapnik, *Support-vector networks*, Machine learning, 20 (1995), pp. 273–297.

[5] N. Cristianini and J. Shawe-Taylor, eds., *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, U.K., 2000.

[6] S. T. Dumais, J. Platt, D. Heckerman, and M. Sahami, *Inductive learning algorithms and representations for text categorization*, in 7th International Conference on Information and Knowledge Discovery, G. Gardarin, ed., 1998, pp. 148–155.

[7] M. C. Ferris and T. S. Munson, *Interior point methods for massive support vector machines*, Technical Report 00-05, Computer Science Department, University Wisconsin, 2000.

[8] A. E. Hoerl and R. W. Kennard, *Ridge regression: Biased estimation for nonorthgonal problems*, Technometrics, 12 (1970), pp. 55–67.

[9] T. Joachims, *Text categorization with support vector machines: Learning with many relevant features*, Lecture notes in computer science, 1398 (1998), pp. 137–142.

[10] ———, *Making large-scale support vector machine learning practical*, in Advances in Kernel Methods, B. Schölkopf, C. Burges, and A. Smola, eds., The MIT Press, 1999, pp. 169–184.

[11] O. L. Mangasarian and D. R. Musicant, *Lagrangian support vector machines*, J. Mach. Learn. Res., 1 (2001), pp. 161–177.

[12] J. C. Platt, *Fast training of support vector machines using sequential minimal optimization*, in Advances in Kernel Methods, B. Schölkopf, C. Burges, and A. Smola, eds., The MIT Press, 1999, pp. 185–208.

[13] M. Pontil and A. Verri, *Support vector machines for 3d object recognition*, IEEE transactions on pattern analysis and machine intelligence, 20 (1998), pp. 637–646.

[14] B. Schölkopf, C. Burges, and V. Vapnik, *Extracting support data for a given task*, in Proceedings, First International Conference on Knowledge Discovery & Data Mining, U. M. Fayyad and R. Uthurusamy, eds., 1995, pp. 252–257.

[15] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. Roy. Statist. Soc. Ser. B, 58 (1996), pp. 267–288.

[16] R. J. Vanderbei, *LOQO: An interior point code for quadratic programming*, Optimization Methods and Software, 11 (1999), pp. 451–484.

[17] V. N. Vapnik, *The nature of statistical learning theory*, Statistics for Engineering and Information Science, Springer-Verlag, New York, 2000.

[18] Y. Yajima, *Linear programming approaches for multicategory support vector machines*, tech. rep., Technical Report 2002–6, Department of Industrial Engineering and Management, Tokyo Institute of Technology, 2002.

[19] Y. Yajima, H. Ohi, and M. Mori, *Extracting feature subspace for kernel based support vector machines*, tech. rep., Technical Report 2001–5, Department of Industrial Engineering and Management, Tokyo Institute of Technology, 2001.