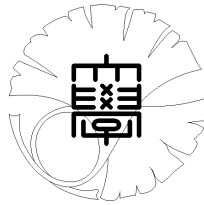


数理科学実践研究レター 2024-6 October 23, 2024

カーネル **SVM** の効率化とヒルベルト **C***加群を使った
カーネル法の応用について

by

向原 未帆



UNIVERSITY OF TOKYO

GRADUATE SCHOOL OF MATHEMATICAL SCIENCES

KOMABA, TOKYO, JAPAN

カーネル SVM の効率化とヒルベルト C^* 加群を使ったカーネル法の応用について

向原未帆¹ (東京大学大学院数理科学研究科)

Miho Mukohara (Graduate School of Mathematical Sciences, The University of Tokyo)

概要

カーネル法は機械学習で非線形データを扱う際に、古典的によく使われる手法の一つである。本研究では、再生核ヒルベルト C^* 加群を使ったカーネル法に関する課題と、カーネル SVM の効率化に関する課題の二つに取り組んだ。その結果、前者に関して、再生核ヒルベルト C^* 加群の Representer 定理 ([1], Theorem 4.8) の成立条件と、近似的な表現定理を得た。また、後者に関しては、ユークリッド空間上の非線形データを使ったカーネル SVM において、データを格子状の点で近似する方法で実行時間を短縮する実験を行った。

1 はじめに

カーネル法とは、与えられたデータを高次元の線形空間に埋め込み、非線形な問題を線形な問題に落とし込む手法のことである。一般に、データ集合 X 上の正定値核 $k: X \times X \rightarrow \mathbb{C}$ に対し、再生核ヒルベルト空間 $\mathcal{H}_k$ と特徴関数 $\phi_k: X \rightarrow \mathcal{H}_k$ が構成される。このような構成により、データを線形空間に埋め込む方法は、機械学習などで有効な手段の一つであり、古くから扱われてきた。

近年、カーネル法の拡張として、複素数体 $\mathbb{C}$ の代わりに C^* 環に値をとる正定値核を使ったカーネル法の機械学習への応用に関する研究が [1] などで行われている。 C^* 環とは行列環や連続関数環の一般化であり、関数解析の分野でよく研究される。 C^* 環を使ったカーネル法では基本的な定理である Representer 定理に特殊な仮定が必要となるが、本研究では仮定の成立条件および、仮定を必要としない近似的な表現定理について調査した。

さらに上記の課題とは別に、カーネル法の効率化に関する実験に取り組んだ。ユークリッド空間上のガウス核を扱うと、データ数の増加に比例して再生核ヒルベルト空間の次元が増加する。線形関数の最適化問題では次元の増加により計算が難しくなることから、データを格子状に射影することで効率化が図れないかと予想し実験を行った。 $\mathbb{R}^2$ 上の二種類のデータを使ったカーネル SVM で実験を行ったところ、適当にパラメータを調整することで実行時間が短縮されたが、パラメータの選び方などに課題の残る結果となった。

2 カーネル法

本節ではカーネル法の導入と基本的な定理の紹介を行う。

定義 1. X を (データ) 集合とする。写像 $k: X \times X \rightarrow \mathbb{C}$ が、正定値核であるとは、すべての $x_1, \dots, x_n \in X$ に対し、 $n \times n$ 行列 $[k(x_i, x_j)]_{1 \leq i, j \leq n}$ が半正定値であることで定義される。正定値核 $k: X \times X \rightarrow \mathbb{C}$ は線形空間 $V_X := \text{span}\{k(x, -) | x \in X\} \subset \text{Map}(X, \mathbb{C})$ 上の内積

$$\left\langle \sum_i c_i k(x_i, -), \sum_j d_j k(y_j, -) \right\rangle := \sum_{i,j} c_i \bar{d}_j k(x_i, y_j) \quad (1)$$

を誘導する。 V_X の完備化を再生核ヒルベルト空間と呼び、 $\mathcal{H}_k$ と書く。

また、写像 $\phi_k: X \ni x \mapsto k(x, -) \in \mathcal{H}_k$ を特徴関数と呼ぶ。各 $\xi \in \mathcal{H}_k$ と $x \in X$ に対し、 $\xi(x) := \langle \xi, \phi_k(x) \rangle$ とすることで $\mathcal{H}_k$ は X 上の関数のなす空間となる。

例 2. 1) $d \in \mathbb{N}$ と $c > 0$ に対し、 $\mathbb{R}^n$ 上の正定値核で $k(x, y) := (x^T y + c)^d$ で定まるものを多項式核という。

2) $t > 0$ に対し、 $\mathbb{R}^n$ 上の正定値核 k で $k(x, y) := \exp(-t\|x - y\|^2)$ で定まるものをガウス核という。

¹miho-mukohara@g.ecc.u-tokyo.ac.jp

カーネル法を使った線形解析で次の定理は基本的である. ([5, Theorem 1.1.2]などを参照.)

定理 3 (Representer 定理). $\mathcal{H}_k$ を再生核ヒルベルト空間, $\phi_k: X \rightarrow \mathcal{H}_k$ を特徴写像とする. 任意の関数 $L: \mathbb{C}^n \rightarrow \mathbb{R}$, 単調増加関数 $g: \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ と $x_1, \dots, x_n \in X$ に対し, $F: \mathcal{H}_k \rightarrow \mathbb{R}$ を $F(\xi) := L((\langle \xi, \phi_k(x_i) \rangle)_{i=1}^n) + g(\langle \xi, \xi \rangle)$ で定める. F が $\mathcal{H}_k$ 上最小値 c をとるとき,

$$c = \min\{F(\xi) \mid \xi \in \text{span}\{\phi_k(x_1), \dots, \phi_k(x_n)\}\}$$

が成立する.

上の定理は $\mathcal{H}_k$ 上の最小化問題が解を持てば, その解が有限次元部分空間の中で実現できることを意味している. これは原理的には単純だが, 応用上有用である.

例 4. ヒルベルト空間 $\mathcal{H}$ 上の有限個のデータ X_0 と X_1 をそれぞれとの距離が最大な超平面 $H(\eta, b) := \{\xi \in \mathcal{H} \mid \text{Re}\langle \xi, \eta \rangle = b\}$ で分離する問題を考える. データが超平面 H で分離できる場合, 適当に $\eta \in \mathcal{H}$ と $b \geq 0$ を選ぶと, 任意の $x \in X_0$ と $y \in X_1$ に対し, $\text{Re}\langle x, \eta \rangle \leq b - 1$, $\text{Re}\langle y, \eta \rangle \geq b + 1$ を満たす. 超平面 $H(\eta, b - 1)$ と $H(\eta, b + 1)$ の距離は $2\|\eta\|$ となるため, 誤分類によるペナルティーを C で定めたときの最適な超平面を決める問題は

$$F(\eta) := \sum_{x \in X_0} \max\{0, \text{Re}\langle x, \eta \rangle - b + 1\} + \sum_{y \in X_1} \max\{0, -\text{Re}\langle y, \eta \rangle + b + 1\} + \frac{1}{C}\|\eta\|$$

を最小化する問題として定式化される. このような最適化問題によってデータを分類する手法をサポートベクトルマシン (SVM) と呼び, 再生核ヒルベルト空間上の SVM をカーネル SVM と呼ぶ. 定理 3 から, カーネル SVM では F の最小解はデータセットの成す部分空間から探せばよい.

3 再生核ヒルベルト C^* 加群

近年, ヒルベルト C^* 加群を使ったカーネル法の機械学習への応用が [1] などで研究されている. 本節では再生核ヒルベルト C^* 加群の Representer 定理について述べる. C^* 環の定義は省略する. [3] などでは基礎的な教科書である.

例 5. 1) 局所コンパクトハウスドルフ空間 X に対し, X 上の無限遠点で 0 に収束する複素数値関数全体の成す環 $C_0(X)$ は C^* 環である.

2) $\mathbb{C}$ 上の行列環 $M_n(\mathbb{C})$ は C^* 環である.

定義 6. A を C^* 環とする. 右 A 加群 $\mathcal{E}$ が半双線形形式 $\langle \cdot, \cdot \rangle: \mathcal{E} \times \mathcal{E} \rightarrow A$ を持ち, 任意の $x, y \in \mathcal{E} \setminus \{0\}$ と $a \in A$ に対し, $\langle x, ya \rangle = \langle x, y \rangle a$, $\langle x, y \rangle^* = \langle y, x \rangle$, $\langle x, x \rangle \geq 0$ を満たすものとする. $\mathcal{E}$ がノルム $\|x\| := \|\langle x, x \rangle\|^{\frac{1}{2}}$ に関して完備であるとき, ヒルベルト A 加群と呼ぶ.

補足. 上記の定義に $\langle x, x \rangle \geq 0$ とあるが, $\langle x, x \rangle$ は C^* 環 A の元である. $a \in A$ がある $b \in A$ により, $a = b^*b$ と書けるとき, a は正である ($a \geq 0$) という. これは A が行列環の時, a が半正定値行列であることに対応する. 本稿では A の正な元全体を A_+ とする. また, $a \geq 0$ かつ $a \neq 0$ の時, $a \succ 0$ とかく.

本稿でヒルベルト C^* 加群の内積は右線形とする. さらに, $\mathcal{L}(\mathcal{E})$ を随伴を持つ作用素 $T: \mathcal{E} \rightarrow \mathcal{E}$ 全体の成す C^* 環とし, $\mathcal{K}(\mathcal{E})$ を $\mathcal{E}$ 上のコンパクト作用素の成す $\mathcal{L}(\mathcal{E})$ の部分 C^* 環とする. (定義は [2] を参照.)

定義 7. C^* 環 A に値をとる写像 $k: X \times X \rightarrow A$ で, 任意の $x_1, \dots, x_n \in X$ に対して, A 上の行列 $[k(x_i, x_j)]_{1 \leq i, j \leq n}$ が正であるとき, k を正定値核と呼ぶ. 定義 1 と同様に, 正定値核 $k: X \times X \rightarrow A$ に対し, X 上の A 値関数の成すヒルベルト A 加群 $\mathcal{M}_k$ と写像 $\phi_k: X \rightarrow \mathcal{M}_k$ が存在し, 次を満たす. 任意の $x, y \in X$ と $\xi \in \mathcal{M}_k$ に対し, $\langle \phi_k(x), \phi_k(y) \rangle = k(x, y)$, $\langle \phi_k(x), \xi \rangle = \xi(x)$. $\mathcal{M}_k$ を再生核ヒルベルト C^* 加群 (RKHM), ϕ_k を特徴写像と呼ぶ.

RKHM の Representer 定理として次が知られている。

定理 8 ([1], Theorem 4.8). $\mathcal{M}_k$ を再生核ヒルベルト A 加群, $\phi_k: X \rightarrow \mathcal{M}_k$ を特徴写像とする. 任意の関数 $L: A^n \rightarrow A_+$, 単調増加関数 $g: A_+ \rightarrow A_+$ と $x_1, \dots, x_n \in X$ に対し, $F: \mathcal{M}_k \rightarrow A_+$ を $F(\xi) := L(\langle \phi_k(x_i), \xi \rangle_{i=1}^n) + g(|\xi|)$ で定める. $\{\phi_k(x_1), \dots, \phi_k(x_n)\}$ の成す $\mathcal{M}_k$ の代数的な部分 A 加群 $\mathcal{N}$ が閉である時, F が $\mathcal{M}_k$ 上最小値 $c \in A_+$ をとれば, ある $\xi \in \mathcal{N}$ で $F(\xi) = c$ となる.

上記の定理では通常の Representer 定理の仮定に加えて, $\{\phi_k(x_1), \dots, \phi_k(x_n)\}$ の成す $\mathcal{M}_k$ の代数的な部分 A 加群 $\mathcal{N}$ が閉であるということが必要である. この仮定は一般には成り立たない.

補題 9. 任意の再生核ヒルベルト A 加群 $\mathcal{M}_k$ に対して, 以下は同値である. (定理 8 の表記を用いる.)

- 1) $\{\phi_k(x_1), \dots, \phi_k(x_n)\}$ の成す $\mathcal{M}_k$ の代数的な部分 A 加群 $\mathcal{N}$ が閉である.
- 2) 射影 $P \in \mathcal{L}(\mathcal{M}_k)$ で像が $\mathcal{N}$ に一致するものが存在する.
- 3) A 上の行列 $a := [k(x_i, x_j)]_{1 \leq i, j \leq n}$ のスペクトルを $\sigma(a)$ とすると, $\sigma(a) \setminus \{0\}$ がコンパクト.

証明の概略 1) と 2) の同値性は [4] の Lemma 2.3.7 から従う. ただし, A が単位的でないときは, 単位化 $\tilde{A}$ を考えればよい. 残りを示す. 一般に, $T: A^n \ni (a_i)_{i=1}^n \mapsto \sum_i \phi_k(x_i) a_i \in \mathcal{M}_k$ はヒルベルト A 加群の間の随伴を持つ作用素となる. $T^*T \in \mathcal{L}(A^n) \cong M_n(A)$ は行列 $a = [k(x_i, x_j)]_{1 \leq i, j \leq n}$ と同一視されるため, $\sigma(T^*T) \setminus \{0\} = \sigma(TT^*) \setminus \{0\}$ から, $\sigma(TT^*) \setminus \{0\}$ がコンパクトであることと 3) は同値である. 1) 及び 2) を仮定すると, 開写像定理から, TT^* が $P\mathcal{L}(\mathcal{M}_k)P \cong \mathcal{L}(\mathcal{N})$ の可逆作用素となるため, $\sigma(TT^*) \setminus \{0\}$ はコンパクトとなる. 逆に $\sigma(TT^*) \setminus \{0\}$ がコンパクトであれば, これは $\sigma(TT^*)$ の開かつ閉な部分集合であるため, 対応するスペクトル射影 P が $\mathcal{L}(\mathcal{M}_k)$ に含まれ, 2) の条件を満たす.

無限次元 C^* 環の元 a に対しては, (連続関数環のような簡単なものでも) 一般には 3) の条件は成立しない. 一方近似的な Representer 定理は次のような形で一般に成立する.

定理 10. $\mathcal{M}_k$ を再生核ヒルベルト A 加群, $\phi_k: X \rightarrow \mathcal{M}_k$ を特徴写像とする. 連続関数 $L: A^n \rightarrow A_+$, 単調増加関数 $g: A_+ \rightarrow A_+$ と $x_1, \dots, x_n \in X$ に対し, $F: \mathcal{M}_k \rightarrow A_+$ を $F(\xi) := L(\langle \phi_k(x_i), \xi \rangle_{i=1}^n) + g(|\xi|)$ で定める. F が $\mathcal{M}_k$ 上最小値 $c \in A_+$ をとれば, 任意の $\epsilon > 0$ に対し, $\{\phi_k(x_1), \dots, \phi_k(x_n)\}$ の生成する代数的な部分 A 加群 $\mathcal{N}$ の元 ξ で $\|F(\xi) - c\| \leq \epsilon$ となるものが存在する.

証明の概略 各ベクトル $\eta_1, \eta_2 \in \mathcal{M}_k$ に対し, $\theta_{\eta_1, \eta_2} \in \mathcal{K}(\mathcal{M}_k)$ を $\theta_{\eta_1, \eta_2}(\zeta) := \eta_1 \langle \eta_2, \zeta \rangle$ で定める. $B_0 := \text{span}\{\theta_{\eta_1, \eta_2} \mid \eta_1, \eta_2 \in \mathcal{N}\}$ は $\mathcal{K}(\mathcal{M}_k)$ の $*$ 部分環で閉包 $B := B_0$ は C^* 環となる. C^* 環は近似的単位元を持つため, B_0 のネット $\{b_i\}_i$ で $0 \leq b_i \leq 1$ を満たし, すべての $\eta_1, \eta_2 \in \mathcal{N}$ に対し, $\lim_i \theta_{\eta_1, \eta_2} b_i = \lim_i b_i \theta_{\eta_1, \eta_2} = \theta_{\eta_1, \eta_2}$ となるものが存在する. 各 $\eta \in \mathcal{N}$ に対して,

$$\|\eta - b_i \eta\|^2 = \|\langle (1 - b_i)\eta, (1 - b_i)\eta \rangle\| = \|(1 - b_i)\theta_{\eta, \eta}(1 - b_i)\|$$

となるため, $\lim_i b_i \eta = \eta$. ここで, $c = F(\xi_0)$ を最小値とする. L の連続性から

$$\lim_i L(\langle \phi_k(x_i), b_i \xi_0 \rangle_{i=1}^n) = \lim_i L(\langle b_i \phi_k(x_i), \xi_0 \rangle_{i=1}^n) = L(\langle \phi_k(x_i), \xi_0 \rangle_{i=1}^n).$$

よって, $\|L(\langle \phi_k(x_i), b_i \xi_0 \rangle_{i=1}^n) - L(\langle \phi_k(x_i), \xi_0 \rangle_{i=1}^n)\| < \epsilon$ となる i が存在する. また $0 \leq b_i^2 \leq b_i \leq 1$ より, $|b_i \xi_0| = \langle b_i^2 \xi_0, \xi_0 \rangle^{\frac{1}{2}} \leq \langle \xi_0, \xi_0 \rangle^{\frac{1}{2}} = |\xi_0|$. 最小性から

$$c = F(\xi_0) \leq F(b_i \xi_0) \leq F(\xi_0) + \epsilon$$

となり, $\|F(b_i \xi) - c\| \leq \epsilon$ を得る. B_0 の定義から $\xi := b_i \xi_0 \in \mathcal{N}$ で, 命題が成立する.

4 カーネル SVM の効率化について

本節では $\mathbb{R}^2$ 上の二種類のデータをカーネル SVM(例 4 で述べた.) で分類する問題を考える. ガウス核を使ったカーネル法では互いに異なる点 $x_1, x_2, \dots, x_l \in \mathbb{R}^2$ に対し, 特徴写像の像 $\phi_k(x_1), \dots, \phi_k(x_l)$ は再生核ヒルベルト空間の一次独立なベクトルとなる. 従って, データ数の増加に比例してヒルベルト空間の次元が増加し, 最適化問題が複雑になる. そこで, $\mathbb{R}^2$ 上のデータを格子上に射影することで, 近似的な分離曲線 (分離超平面の引き戻し) を求める実験を行った. 実験には scikit-learn を使った.

4.1 実験内容

図 1 のような茶色と水色のデータを N 分割し, それぞれ $\Gamma_l := \frac{1}{l}\mathbb{Z}^2$ 格子上の最も近い点に射影する (図 2). これらの和をとると, Γ_l の 2 色のデータで, 同じ格子点上に同じ色のデータが高々 N 個であるものを得る. これらを使い, 以下の実験を行った.

- 1) 元の学習データから得た分離曲線と, 格子上のデータから得た分離曲線の視覚的な比較.
- 2) 元の学習データを使ったテストデータの分類と格子上の学習データを使った分類の, 精度と実行時間の比較.

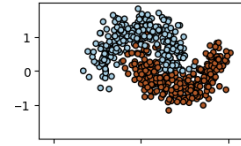


図 1. 元のデータ

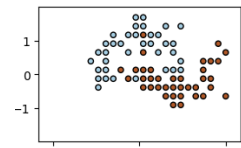


図 2. Γ_l 上のデータ

4.2 結果

- 1) データ数 $n = 500$, 分割数 $N = 5$ を固定, 格子サイズ $\frac{1}{l}$ を変化させたところ, 分離曲線 (図 3, 4, 5 の実線) は l の増加に伴い, 元のデータから得られるもの (図 6 の実線) に近づいた. この実験では, 例 4 にあるように, 茶色のデータ (X_0) と水色のデータ (X_1) を再生核ヒルベルト空間 $\mathcal{H}_k$ 上の超平面で分離した. 以下の各図内の 2 本の点線は $\mathcal{H}_k$ 上で距離が最大となる平行な分離超平面の組 (例 4 の $H(\eta, b-1)$ と $H(\eta, b+1)$) の引き戻しを表す. また, 各図の実線は中間の分離超平面 (例 4 の $H(\eta, b)$) の引き戻しである.



図 3. $l=1$



図 4. $l=1.5$

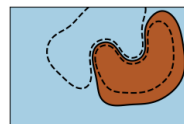


図 5. $l=2$

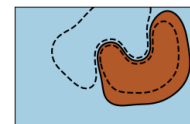


図 6. 元データ

- 2) 格子サイズ $\frac{1}{l} = \frac{1}{5}$ と分割数 $N = 5$ を固定し, 学習データ数 n を変化させた場合, テストデータの分類の精度と実行時間に関して, 以下の結果を得た.

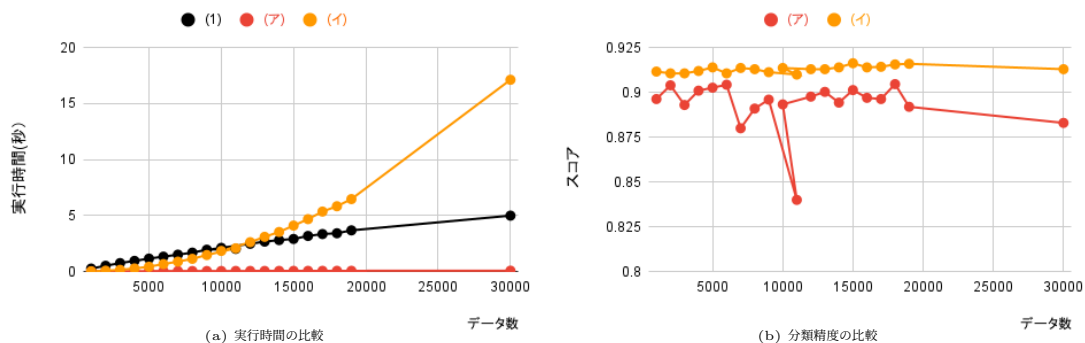


図 7. (ア) は格子上の学習データによる分類の精度と実行時間を, (イ) は元の学習データによる分類の結果を表す. また, (1) はデータの格子上的射影にかかった時間を表す.

グラフ (a) の縦軸はカーネル SVM の実行時間を, グラフ (b) の縦軸はカーネル SVM でテストデータが正しく分類された割合を示している. データ数が十分大きい場合, 格子への射影により実行時間が短縮された. しかし, 今回の実験では 1000 程度の学習データで, 短時間で高い精度の分類が行われたため, 有用性の検証はできなかった.

5 終わりに

本稿では, 機械学習等で頻繁に扱われるカーネル法に着目し, 近年の研究 [1] に関する考察と, 古典的なカーネル SVM の効率化に関する実験を行った. 第 3 節では, [1] の結果である再生核ヒルベルト C^* 加群の Representer 定理について調査し, 定理の仮定の成立条件と, 一般の状況での近似的な結果について述べた. 仮定が成立する非自明な正定値核の例や, 近似的な結果の応用例なども今後考えたい. 第 4 節では, カーネル法の効率化を目指し, 実験を行った. $\mathbb{R}^2$ 上のデータよりも格子点上のデータの方が再生核ヒルベルト空間の次元が落ちるため, 最適化問題が簡単になると予想したが, 実際には, 格子点で近似するだけでは実行時間の差はほとんど得られなかった. 各格子点上のデータの数を高々 N に制限することで上記の結果を得たが, この方法ではデータの種類やノイズの大きさによって N や l をうまく選ぶ必要がある. また, データ数が 1000 程度では実行時間はほとんど変わらなかったため, 有用性の検証にも至らなかった. 大量の学習データが必要になるケースでどうなるのか, 分割数や格子サイズをどのように選べばよいか, より詳しい検証が必要であると考えている.

謝辞 本研究にあたり, 課題のご提案及び, 多くのご助言をくださった株式会社ニコンの高山侑也氏, 信田萌伽氏に感謝する. また, 研究とセミナー遂行のサポートを頂いた東京大学数理科学研究科の野村亮介氏, 鮑園園氏, 森迪也氏, 小原和馬氏, 論文 [1] の内容について解説をしてくださった NTT ネットワークサービスシステム研究所の橋本悠香氏, 本稿を査読し改善の提案をくださった査読者の方に感謝する.

参考文献

- [1] Y. Hashimoto, I. Ishikawa, M. Ikeda, F. Komura, T. Katsura, Y. Kawahara; *Reproducing kernel Hilbert C^* -module and kernel mean embeddings*, JMLR **22**, 267:1–56, 2021.
- [2] E. C. Lance; *Hilbert C^* -modules – a toolkit for operator algebraists*, volume 210 of London Mathematical Society Lecture Note Series. Cambridge University Press, England, 1995.
- [3] G. J. Murphy; *C^* -Algebras and Operator Theory*, Academic Press, Inc., New York, 1990.
- [4] V. M. Manuilov, E. V. Troitsky; *Hilbert C^* -and W^* -modules and their morphisms*, Journal of Mathematical Sciences **98**, No. 2, 137–201, 2000.
- [5] G. Wahba; *Support vector machines, reproducing kernel Hilbert spaces and the randomized GACV*. Advances in Kernel Methods - Support Vector Learning, pages 69–88, Cambridge, MA, 1999. MIT Press.