はじめに

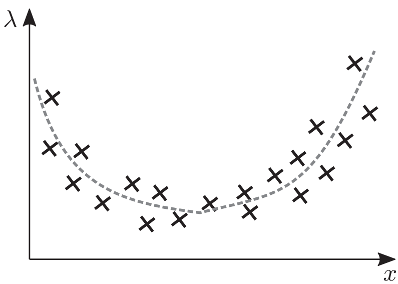
近年，人工知能に関連して機械学習という言葉が広く知られるようになった．本書は，機械学習の背景にある関数解析の入門書である．広範に及ぶ機械学習の定義付けは専門書に譲ることとして，本書では変数 $\boldsymbol{x} \in \mathbb{R}^d$ および $\lambda \in \mathbb{R}$ の計測データ $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^d$ と $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ が与えられたとき，データを元に $\boldsymbol{x}$ と λ の関係を近似的に推定することを**学習**とよぶ．特に，本書では回帰問題と分類問題という二つの問題を扱う．

回帰問題

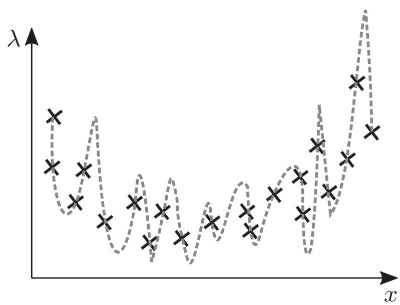
計測データ $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ と $\lambda_1, \ldots, \lambda_n$ を用いて，変数 $\boldsymbol{x}$ および λ との間に，誤差を許して

$$\lambda = f(\boldsymbol{x}) \tag{0.0.1}$$

という関係を当てはめることを考えよう．例えば，$d = 1$ としてデータ (x_i, λ_i) を**図 0.1** (a) のように ‘×’ 印で示すとき，すべてのデータがその関数の周辺に分布する，破線で示すような関数 f を求めるのである．この問題は**回帰問題**



(a)　関数 f の一例



(b)　過学習の一例

図 0.1：回帰問題の例

とよばれる．例を示そう．一般に，一日の消費電力のピーク値 λ [W] はその日の最高気温 x [°C] に依存することが知られている．今，天気予報から明日の最高気温の予報値 x_* が得られたとき，明日の消費電力ピーク値 λ_* を予測したいとする．ここで，過去 n 日分の最高気温データ $x_1, \ldots, x_n$ と対応する消費電力ピーク値のデータ $\lambda_1, \ldots, \lambda_n$ が与えられるとする．このとき，これらのデータを用いて，x と λ を関係付ける関数 f が求まれば，

$$\lambda_* = f(x_*)$$

から明日の消費電力ピーク値 λ_* を予測することができるであろう．回帰問題の解を求める作業は**学習**の一例である．データが更新される度に繰り返し学習することで予測精度の向上が期待される．

さて，データの数は有限であるから，図 0.1 (b) に示すように，複雑な関数を使えばすべてのデータ $(x_1, \lambda_1), \ldots, (x_n, \lambda_n)$ が完全に曲線 (0.0.1) 上にのる，すなわち

$$\lambda_i = f(x_i) \quad (i = 1, \ldots, n)$$

をみたすような f は存在する．しかし，それは当初のデータに強く依存したものかもしれない．例えば，データが更新される度に全然異なる関数が出てくることも考えられる．もしそうなってしまったら，本来の目的である予測のためには相応しくない．このような状況を**過学習**という．過学習を避けるために，関数を簡単なものに制限する．仮に簡単な関数とは一次関数のことであるとすれば，この問題は簡単に解くことができるが，事態は単純ではない．例えば，消費電力のピーク値は気温の高い夏と気温の低い冬に高まることが知られており，一般にデータは図 0.1 (a) に示すようにおわん型の関数の周辺に分布する．このような場合に x と λ を一次関数 f で関係付けるのは不適切であろう．

分類問題

次に，$D = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ とし，各データ $\boldsymbol{x}_j$ には符号 $\lambda_j \in \{-1, +1\}$ がラベル付けされているとする．例えば，健康診断を想定すると，d 個の検査項目（腹囲，BMI, 中性脂肪など）からなる n 人分の健康診断結果があり，さらに，

検査から 10 年後の生活習慣病発症の有無が符号 ± 1 で記録されている状況を考えている．ここで，D の分割

$$D_+ = \{\boldsymbol{x}_j \in D : \lambda_j = +1\}, \quad D_- = \{\boldsymbol{x}_j \in D : \lambda_j = -1\}$$

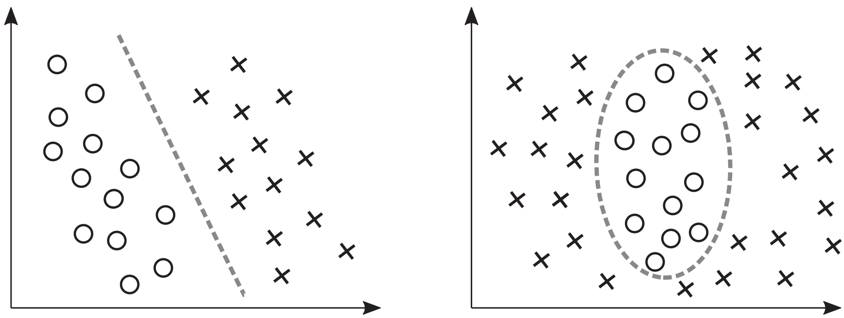
に対し，

$$D_+ \subset \{\boldsymbol{x} \in \mathbb{R}^d : f(\boldsymbol{x}) > 0\}, \quad D_- \subset \{\boldsymbol{x} \in \mathbb{R}^d : f(\boldsymbol{x}) < 0\}$$

をみたす簡単な関数 f を見つけたい．このような問題は**分類問題**とよばれる．これができれば，まだラベル付けされていない新しいデータ $\boldsymbol{x}$ に対し $f(\boldsymbol{x})$ を求めて，$f(\boldsymbol{x})$ の符号により $\boldsymbol{x}$ のラベルの予測，すなわち将来の病気の予測を与えることができる．分類問題の解を求める作業も**学習**の一例である．回帰問題と同様に，データが更新される度に繰り返し学習することで予測精度の向上が期待される．さて，データの数は有限であるから，複雑な関数を使えば D_+ と D_- は必ず分離できるが，ここにもやはり過学習の問題が生じる．そこで f を一次関数であるとすれば，この問題は

「ユークリッド空間内に分布している点を超平面で分割できるか？」

と翻訳することができる．今，平面上のデータ D_+ と D_- をそれぞれ '$\circ$' 印および '$\times$' 印で描画するとき，これらが**図 0.2** (a) のように分布しているの



(a)　一次関数 f で分類できる場合　　**(b)**　一次関数 f で分類できない場合

図 0.2：分類問題の例

であれば，一次関数 f による分離は可能である．ところが，データの分布がユークリッド空間の超平面と相性が良いとは限らない．例えば，データが図 0.2 (b) のように分布する場合にはどのように一次関数 f をとっても D_+ と D_- を分離することはできない．一次関数では単純すぎるのである．

カーネル法とは

回帰問題と分類問題にはともに，一次関数では単純すぎるが，複雑な関数を使うと過学習の問題が現れた．そこで当初の回帰問題，分類問題そのものを代入が内積で表される空間での問題に変換し，その変換した先での一次関数を考える．それが**カーネル法**である．具体的には，**カーネル関数**とよばれる関数 $k(\boldsymbol{x}, \boldsymbol{y})$ を与え，データ $\boldsymbol{x}_j$ を関数 $k_{\boldsymbol{x}_j}(\boldsymbol{x}) = k(\boldsymbol{x}, \boldsymbol{x}_j)$ に変換する．これを**カーネルトリック**とよぶ．例えば，3 つの地点 A, B, C 間の距離や角度の情報を知りたいとする [*1]．通常，我々は地点 A, B, C を地図上の位置と対応付けてから，これらの情報を知る．すなわち，それぞれの地点の地図上の位置を与える写像 Φ を用いて，$\Phi(A)$, $\Phi(B)$, $\Phi(C)$ 間の距離や角度を A, B, C 間の距離や角度とするのである．距離や角度は内積によって測られることに注意すれば，距離や角度を考えることは $\{A, B, C\}$ に内積を与えることに相当する．第 4 章で詳しく解説するが，この考えを一般化すればカーネル関数が自然に現れる．実際，$\langle \cdot, \cdot \rangle_{\mathbb{R}^2}$ を $\mathbb{R}^2$ の通常の内積とすれば，$\Phi : \{A, B, C\} \to \mathbb{R}^2$ がどのような変換であっても

$$k(x, y) = \langle \Phi(x), \Phi(y) \rangle_{\mathbb{R}^2} \quad (x, y \in \{A, B, C\})$$

は $\{A, B, C\}$ 上のカーネル関数である．カーネル関数を取り換えれば，回帰問題，分類問題の解を様々な関数から選ぶことができる．さらに，本書ではほとんど触れることができないが，カーネルトリックにより計算量が増えることもない．これもカーネル法の特徴である．

[*1] この例えは荷見守助先生（茨城大学名誉教授）にアイデアをいただいたものである．

本書の構成

本書では理工系学部の標準的な数学の知識を前提に「機械学習のための関数解析入門」と題してカーネル法の理論と応用の解説を試みる．第 1 章では内積の計算を中心に線形代数の復習をしよう．第 2 章では，フーリエ解析と複素解析からいくつかの事実を認めて，内積の数学としてのフーリエ解析を解説する．第 3 章ではヒルベルト空間の基礎理論を解説する．ヒルベルト空間とは，第 1 章と第 2 章の数学に共通した構造を抽出した概念である．ここで抽象的な内積の計算に慣れてしまえば，カーネル法の理解は難しいことではない．第 4 章ではカーネル法の基礎を，理論と応用を交えて解説する．第 5 章ではカーネル法の発展編としてガウス過程回帰を解説する．ここで数学の枠を超えた本格的な応用を紹介しよう．付録では，本書を読む上で知っておくと便利なことや，少々進んだ話題をまとめた．さて，本書の読み方であるが，目的や事前の知識の量に依り，様々な道筋が考えられる．例えば，カーネル法を手短に知りたい場合，第 1 章から第 4 章，第 5 章と進むことが可能であろう．また，半期の講義で使用する際は，第 1 章，第 3 章，第 4 章を中心としたコースが適当であろう．数学や情報科学専攻の学生には，卒業研究などでの通読を勧めたい．

機械学習は工学と数学の境界に位置するため，本書執筆中に著者自身が新たに学んだことは多い．そのため構想は今も広がり続けているのだが，本書全体のまとまりを考えて今回は見送ったテーマがいくつかある．例えば，第 4 章と第 5 章にていくつかの応用は紹介したものの，本書だけでは実践的な応用力は身につかないだろう．また，機械学習は人工知能との関連で世間に注目されたが，システム制御をはじめとする工学一般での応用を外すことはできない．今回扱うことができなかったこれらのテーマについては，応用編として近い将来にまとめたい．

最後に，本書の完成までに多くの方々のお世話になりました．本書執筆のきっかけとなる機会を提供してくれたのは広島工業大学の谷口哲至氏です．伊吹賢一氏と防衛大学校の土田兼治氏は初期の原稿を精読し，多くの誤りや読みづらさを指摘し，改善を提案してくれました．茨城大学名誉教授の荷見守助先生に

は，原稿だけでなく，本書の計画全体について継続的な助言と励ましをいただきました．そして，本書の出版にあたっては，内田老鶴圃社長内田学氏，同社編集部笠井千代樹氏，生天目悠也氏に大変お世話になりました．皆様に厚くお礼を申し上げます．

2021 年 2 月

著者