

Github 版

一般的^{*1}な機械学習入門

最終更新：2020 年 10 月 11 日

逢空れい@ranoiaru

^{*1} ここで言う「一般的な」とは、「普通の人にもわかりやすい」という意味ではなく、数学界限における「汎用性の高い」「広く対応可能な定義を導入している」という意味である。

目次

1	はじめに	3
1.1	本書執筆の経緯（よまなくてもいいよ）	3
1.2	前提知識	3
2	入門：機械学習の一般的問題設定	4
2.1	頻度論的機械学習問題の定義	4
2.2	ベイズ論的機械学習問題の定義	4

notation

- \mathcal{X} : 特徴量空間。 \mathbb{R}^d の単連結な開部分集合
- \mathcal{Y} : ラベル空間。 \mathbb{R}^n もしくは \mathbb{C}^n
- (Ω, \mathcal{F}, P) : 確率空間
- H : 使う手法により異なる条件を満たす関数 $f : \mathcal{X} \rightarrow \mathcal{Y}$ の集合が為す可分ヒルベルト空間
- L : 頻度論的手法における損失関数 $H \times \Omega \rightarrow \mathbb{R}$
- $\nabla_f L(f, \omega), f \in H$: 頻度論的手法における勾配の定義。 ω を固定した上でのフレシェ微分に f を代入したもの。
- $\mathcal{B}(A)$, A は距離空間 : ボレル集合族
- $\mathcal{P}((\Omega, \mathcal{F}))$: 可測空間 (Ω, \mathcal{F}) 上の確率測度全体の集合。 \mathcal{F} を略記し、 Ω が距離空間である場合、 $\mathcal{P}(\Omega)$ は $\mathcal{P}((\Omega, \mathcal{B}(\Omega)))$ であるものとする。
- \mathcal{R} : リッジレット作用素
- \mathcal{R}^* : 双対リッジレット作用素
- D_{ucp} : ucp 位相の入った càdlàg な適合過程の集合。(すなわち発展的可測)
- L_{ucp} : 同上。ただし上が確率積分の定義域であるのに対して、こちらは値域。確率過程として動いている空間が違ったりする。
- N : ポアソンランダム測度

1 はじめに

深夜テンションによる出来心で「数学徒向けの機械学習入門書を書いたら見てくれる人ふぁぼください。ふぁぼ多かったら実行します」と書いたら、なんか一晩で 500 以上ふぁぼられたうえ、なぜか 200 近くフォロワーが増えたのでやらざるを得ない。

1.1 本書執筆の経緯（よまなくてもいいよ）

本書の構想は三年前に遡る。当時私は機械学習に入門したばかりの修士課程学生だったが、その時に触れた書籍があまりに数学的厳密性にも一般性にも欠け、そのことに対する苛立ちをツイートにぶつけた。

その結果、燃えた。

クソリブと暴言の嵐で通知が埋まり、私はしばらく鍵垢に籠った。

私の言い方も大いに悪かったのだが、あまりにも曲解されすぎではないかと当時の私は思ったものである。あの発言の意図としては、インターン時のパワハラ面接官や、twitter のひたすらマウントをとってくる某エンジニアに対して「数学ろくに知らないくせに『数学なんて機械学習に無意味』なんていうんじゃない」ということであり、断じて「純粋数学に明るくなきゃ本当に機械学習わかってるとは言えない」などということではない。誓ってもいい。

その時私は、数学徒向けの機械学習入門理論を整備すると宣言し、実際に修論にはそれを書いたり、数学徒のついでに講演したりしたのだが、このように資料の形での公開はほとんどしていない。

機械学習界限を騒がせてからちょうど三年。私も学生時代が終わり、数学寄りの機械学習研究者として仕事をしながら、修士時代の研究室へ社会人博士への入学をもくろむ日々。ツイートがバズったのもなにかの縁であろう。変な書き方のせいで不快な思いをさせた方々へのお詫びも兼ねて、ここはひとつ、数学徒向け機械学習入門書の執筆を行うことにした。

1.2 前提知識

本資料において、入門編の前提知識は「学部レベルの解析学」（主に関数解析とルベーグ積分）、発展編はその都度必要な知識を補完されたし。

確率解析は基本的に [1] の流儀を使用する。^{*2}

^{*2} 数か月前まで無駄に難解なだけの読みにくい本だと思っていましたが、変な確率過程の研究をしたらこんなにありがたい既存理論は他にないと分かります。この本がなかったら、私がここ 4 か月で出したの成果と同じ研究成果を出すのに 3 年くらいかかってそう。「ucp 位相の入った適合 *cádlág* 過程空間」の確率解析理論がそこそこ完成されてる事実があまりに便利すぎる。

2 入門：機械学習の一般的問題設定

この章では、機械学習の問題設定について、関数解析的に定義する。

定義 2.1 データの定義

多数の特徴量とラベルの組、 $\{(x_i, y_i)\}_i \subset \mathcal{X} \times \mathcal{Y}$ をデータと呼ぶ。

2.1 頻度論的機械学習問題の定義

2.1.1 仮説空間

特徴量空間からラベル空間への写像のうち、手法によってことなる条件を満たすものが為す可分ヒルベルト空間 H を仮説空間と呼ぶ。この H の中から、なるべく良い f を選ぶのが、機械学習問題の目的である。

2.1.2 損失関数

損失関数 L と仮説空間 H の組は次を満たす必要がある。

定義 2.2 許容可能な組み合わせ

任意の $f \in H$ に対して、フレシェ微分に f を代入した値 $\nabla_f L(f)$ が H に含まれる。^{*3}

次の定理は関数解析的定義と、実際のアルゴリズムを結ぶ超重要定理である。

定理 2.1 フレシェ微分はユークリッド空間における勾配の拡張

H が有限次元ユークリッド空間 \mathbb{R}^N と同型であるとき、フレシェ微分はユークリッド空間の勾配 $\nabla_\theta, \theta \in \mathbb{R}^N$ と一致する。

2.2 ベイズ論的機械学習問題の定義

^{*3} 機械学習がある程度分かる方はこの定義を「勾配法で更新した結果、モデルの枠組みをはみ出した関数になったら困るから」と理解していただければわかりやすいでしょう。もうちょっと拡張できますが、めんどくさすぎるし、普通に使われてる実アルゴリズムは全部この条件を満たすのでこれでいきます。

参考文献

- [1] P. Protter, Stochastic Integration and Differential Equations, Applications of Mathematics, Second edition, Vol. 21 (Springer-Verlag, Berlin, 2005).