

機械学習原論

2018 年 2 月 12 日

1 はじめに

世は空前の人工知能ブーム。猫も杓子も AI だディープラーニングだと、なんでも機械学習に任せてしまえばいいという考えが蔓延している。

機械学習エンジニアの待遇の良さや需要の高さを耳にする機会は多いことだろう。かつて数学徒の目指す職種としてありがちだったクオন্ツやアクチュアリーと違い、スーツを着ることを強要されるのが少ないという点も魅力的だ。そうなる「自分も機械学習を学んで AI エンジニアになり、ガッポガッポ大儲け。タワーマンションの上層階でたくさん女侍らして、高級ワインを飲むんじゃあ！」という思考になり、機械学習の本に手が伸びる数学徒も少なくないことかと思う。

しかし、勉強をし始めた多くの数学徒は匙を投げてしまう。その理由の大半が「あまりに数学的に適当すぎて読んでいられない」というものだ。

その気持ちはよくわかる。かく言う私もそうだった。世間には「一般向け」と称した、あまりにも粗雑な機械学習の入門書が溢れかえっており、数学的にまともな文献は少数派だ。もちろん一部存在はするが、とてもじゃないが初学者が読んで役立つことはあまりに難しい。

なぜ読むのが難しいのか。理由としては、大きく分けて 3 つあるように思う。

1. 確率測度と確率密度関数が区別されていない
2. 確率変数であるものと確率変数でないものが区別されていない
3. 目の前の関数がどこからどこへの写像かわからない

慣れてくるとなんとなく忖度して雰囲気を読んでいくことができるようになるが、厳密に読んでいく訓練を受けてきた数学徒にはなかなか酷な話だ。

そのため、本書の前半ではなるべく数学科の解析学の授業で習う用語、流儀を用いて、機械学習の基礎を書くことを心掛けた。本書の内容を理解すれば、そう苦労せず一般向けの機械学習資料を読んでいくことができるはずだ。

前半部分の前提知識は数学科学部レベルの関数解析学、確率論、数理統計学、微分方程式論とする。

後半では数学の知見を用いた機械学習研究について解説する。こちらでは普通に院レベル以上の解析学を用いる。

2 機械学習入門

2.1 問題設定

特に表記がなければ、確率空間 (Ω, \mathcal{F}, P) の σ 加法族はボレル集合族、線形空間に対応する体は実数体であるものとする。

線形位相空間 \mathcal{X} を特徴量空間、線形位相空間 \mathcal{Y} をラベル空間と呼び、それぞれの元 x, y を「特徴量」「ラベル」と呼ぶ。

ここで、独立同分布な確率変数列 $\mathcal{D} := \{X_i(\omega)\}_{i=1}^n$ を考える。各 X_i は $X_i : \Omega \rightarrow \mathcal{X} \times \mathcal{Y}$ となる可測関数である。この確率変数列 $\mathcal{D} (= \{X_i\}_{i=1}^n)$ を入力データと呼ぶ。

増大情報系 $\mathcal{F}_0 \subset \mathcal{F}_1 (\subset \mathcal{F})$ を考える。 \mathcal{F}_0 はデータ観測前の保持情報、 \mathcal{F}_1 はデータ観測後の保持情報である。すなわち

$$\mathcal{F}_1 := \mathcal{F}_0 \vee \sigma([X_i]_{i=1}^n) \quad (1)$$

連続写像 $f : \mathcal{X} \rightarrow \mathcal{Y}$ の満たす集合全体を Y と表記する。モデルに対応した Y の部分集合 (詳細は後述) を Y_M と書く。

汎関数 $F : Y \rightarrow \mathbb{R}$ の集合を Y^* と書く。位相は弱位相であるとし、 \mathcal{F}_1 -可測な確率変数 $F^* : \Omega \rightarrow Y^*$ によって定義される $F_{\mathcal{D}} := F^*(\omega)$ に対して

$$\hat{f} := \operatorname{argmin}_{f \in Y_M} F_{\mathcal{D}}(f) \quad (2)$$

を満たす $\hat{f} \in Y_M$ を見つける。これが機械学習における大枠の問題設定である。

参考文献

- [1] 確率論 (実教理工学全書), 西尾真紀子 (1978)
- [2] Machine Learning: A Probabilistic Perspective (Adaptive Computation and Machine Learning series), Kevin.P.Murphy (2012)
- [3] ルベーク積分によう入門-使うための基礎と応用-, 吉田伸生 (2006)