機械学習の数理

宮本来夏:@ranoiaru @riranoiaru

2018 年 9 月 4 日

1　はじめに

前回、「来年のつどい講演テーマは数理医学」と言ったな、あれは嘘だ。

今回のテーマは、数理ファイナンスが続いていた前回までとは打って変わって「機械学習」である。

2　講演内容

2.1　機械学習入門 (30 分程度の予定)

機械学習入門書を投げ出してしまう数学関係者は多い。

というのも、あまりに数学的に適当な入門資料が世の中に氾濫しているからだ。ある程度分かってくれば、なんとなく忖度して読んでいくことができるようになるが、厳密に読んでいくことを心掛けてきた機械学習初学者の数学徒には酷な話である。

そこで、この講座では、実用上深刻な不便を起こさない範囲でなるべく純粋数学の流儀を用いて機械学習において使われる notation、および機械学習という応用数学の概要について解説する。「純粋数学のお気持ち」から「機械学習のお気持ち」への橋渡しとなる講座にしたい。

内容としては、2018/3/4 に行われた阪ゼミ講演会での講演内容を大幅に改良し、コンパクトにまとめたものとなる。

2.2　ニューラルネットワーク (入門講座を終えた後の残り時間の限り)

入門講座の後は、近年世間を騒がせている「深層学習」について解説する。

深層学習とは、「ニューラルネットワーク」と呼ばれる計算手法のうち、4 層以上で構成されているものを指し、現在の機械学習研究の主流派である。

2.2.1　積分表現理論

いわゆる wide continue である。3 層のニューラルネットの中間層を無限大にした場合を考える。

近似したい関数を $f \in L^2$ とし、リッジレット変換と呼ばれる作用素 $\mathcal{R}_\rho$ を用いると、適当な条件の下で

$$\int_{\mathbb{R}^d \times \mathbb{R}} (\mathcal{R}_\rho f)(a,b)\sigma(a \cdot x - b)d\mu(a,b) = f(x) \tag{1}$$

が f の定義域上 $a.e.$ で成立する。すなわち、リッジレット変換を計算することで、最適なパラメータを一

発の数値計算で求めることができる。

2.2.2　輸送解析

いわゆる depth continue である。ニューラルネットを無限に深くしていくとどうなるかの解析だ。

$\mathcal{X} := \mathbb{R}^d$ 上の確率測度のうち、一定の条件を満たすものを $\mathcal{P}(\mathcal{X})$ と置き、その上での距離 W_2 を

$$W_2(\mu, \nu) := inf_{\pi \in \Pi(\mu,\nu)}[\int_{\mathcal{X}^2} c(x, y)\pi(dx, dy)] \tag{2}$$

と置く。すると $(\mathcal{P}(\mathcal{X}), W_2)$ は適切に定義された接ベクトル空間とリーマン計量によって、リーマン多様体となる。

このリーマン多様体上の勾配流として記述される偏微分方程式が、深層ニューラルネットの連続化と同等の存在である（≒深層ニューラルネットとはその輸送の離散化である）という議論についてお話する。

2.2.3　残差学習と微分方程式

これは depth continue 的思考の別形態であると言える。

近年、CNN を著しく深くするネットワークの構築法として、ResNet と呼ばれる手法が流行りである。

これは、それまでの深層学習ではあるブロックで学習させたい写像 $H : \mathbb{R}^d \to \mathbb{R}^d$ を直接学習させていた。

しかし、残差学習と呼ばれる手法では $F(x) := H(x) - x$ を学習させ、後から x を足し合わせる。

n ブロック目の入力を x_n と置くと

$$x_{n+1} = x_n + F_n(x_n) \tag{3}$$

となり、これは常微分方程式のオイラー法による近似に他ならない。

すなわち、ResNet とは、常微分方程式

$$\partial x_t / \partial t = F(t, x_t) \tag{4}$$

という常微分方程式を、探索で解く機構であると言える。

こう考えることにより、圧倒的に長い歴史を持つ微分方程式論を援用した深層ニューラルネットの解析が可能となる。

3　前提知識

数学科学部レベルの微分幾何学、関数解析学、微分方程式論、確率論、統計学。

以下の資料に目を通しておいていただけると嬉しい。

参考文献

[1] 深層ニューラルネットの積分表現理論, 園田翔 (2017)

[2] Beyond finite layer neuralnetworks:bridging deep architectures and numerical differential equations,Yiping Lu,Aoxiao Zhong,Quanzheng Li(2017)

[3] Integral representation of the global minimizer,Sho Sonoda(2018)