数学から見た機械学習，機械学習から見た数学

monariz (Twitter:@monariz_thmap)

2021/03/20・21

1　概要

最近、数学科卒の方がドンドン機械学習へ進出しているのを肌で感じます。何を隠そう、私もその 1 人です。これは主観ですが、大学で学んだ数学を活かせる職業は今までアクチュアリーやクオンツ、研究者に限られてきたと感じています。*1 しかし、最近第 3 次 AI ブームという波が来ており（そろそろ落ち着くともいわれていますが。)、「機械学習エンジニア」がその選択肢に 1 つに加わろうとしています。数学科卒から機械学習に進んだ人間として、数学と機械学習の架け橋になれれば、と思う次第でこのタイトルを設定しました。

2　第一部　数学からみた機械学習 (60 分)

第一部では、現在の機械学習、特に深層学習（ディープラーニング）を支える重要な定理「ニューラルネットワークの普遍性定理」を紹介し、証明します。

Th 1　ニューラルネットワークの普遍性定理 (G.Cybenko 1989) [1]

連続関数 $\sigma(t)$ を

$$\sigma(t) \to \begin{cases} 1 & as\ t \to +\infty \\ 0 & as\ t \to -\infty \end{cases} \tag{1}$$

を満たすものとする。また、十分大きい N に対して、関数 $g : \mathbf{R^n} \to \mathbf{R}$ を

$$g(x) = \sum_{j=1}^{N} \alpha_j \sigma(y_j^T x + \theta_j)$$

と定義する。($y \in \mathbf{R^n}, \alpha, \theta \in \mathbf{R}$) このときこの形をした関数全体の集合 $G = \{g(x) | \alpha_j \in \mathbf{R}, \mathbf{y_j} \in \mathbf{R^n}, \theta_\mathbf{j} \in \mathbf{R}\}$ は、n 変数連続関数全体集合のなかで稠密。

*1 他にもありましたらごめんなさい

言い換えれば、任意の正の数 ϵ と、任意の連続関数 $f : \mathbf{R^n} \to \mathbf{R}$ に対して、ある $N, \alpha, y, \theta, g(x) \in G$ が存在して、

$$\forall x \in \mathbf{R^n},\ |\mathbf{f}(\mathbf{x}) - \mathbf{g}(\mathbf{x})| < \epsilon$$

が成り立つ。

この定理がなければ、現在の第 3 次 AI ブームがなかったであろうと呼ばれる定理です。当日はこの証明および、利用法についてお伝えします。

3 第二部 機械学習から見た数学 (30 分)

概要でもお伝えしました通り、機械学習には大学で習う数学が必要です。「え、どうせ微分積分と線形代数、統計だけじゃないの？」と思う方もいらっしゃるかもしれません。たしかに、機械学習のベース理解に必要な知識はそれらだけですが、実は数学科でしか習わないような、学部数学レベルも論文を読んでいくとちょこちょこ出てきます。（関数解析はもとより、群論、多様体や双曲空間など）それらをかいつまんでお話しできればと思います。

4 対象者・予備知識

対象者:学部 1 回生より
予備知識:教養レベルの微分積分が分かれば雰囲気は掴めます。

参考文献

[1] G.Cybenko (1989) "Approximation by superpositions of a sigmoidal function", Mathematics of Control, Signals, and Systems, 2(4), 303–314

[2] その他、紹介する論文は当日