

論文内容の要旨

論文題目 アンサンブル学習とベイズ法の予測性能の解析

氏名 伏木 忠義

既存のデータをもとに将来のデータの値を予測する方法は、日常で広く必要とされる技術である。このような予測問題に対して、統計学や学習理論など、さまざまな分野で研究が行われてきた。予測問題は基本的で重要な問題であり、長年の研究があるが、使える計算リソースや扱う問題の大きさ、データ量といった要因は、時代と共に変化し、それらに対応した手法が提案、研究されている。

最近の学習理論の世界では、アンサンブル学習とよばれる手法が研究されている。アンサンブル学習は、何度も学習を行い、その結果得られた多数の学習機械をうまく組み合わせて予測を行うものである。アンサンブルをとることで1つの学習機械のみで予測するよりも良い予測が実現される場合がある。特に、ブートストラップデータを用いて何度も学習を行い、それらの単純平均で予測する bagging や学習がうまくいかなかった例題について重みをつけて次の学習を行う boosting という手法が有名である。

一方、統計学においては、統計的予測問題の文脈で、ブートストラップデータを用いて予測する手法や Bayes 法を用いた予測法が研究されている。統計学で提案されていたブートストラップを用いた予測は、学習理論のアンサンブル学習の立場からは、統計的予測問題に bagging を適用したものと考えることができる。Bayes 予測は、事後分布からパラメータをリサンプリングして平均をとったものと考えられるので、これもアンサンブル学習の1つとみなせる。

本論文では、このように2つの異なる分野で研究されていた手法を統一的な観点でとらえ、Bayes 的な立場から統計的予測問題におけるアンサンブル学習について調べる。情報幾何学の枠組みを用いることで見通しよく議論を進めることができるようになる。

まず、ブートストラップを用いた予測は、ある条件のもとで、Bayes 予測の近似となっていることを示す。幾何学的には、ブートストラップを用いた予測分布は、最尤推定量を用いた予測分布にモデルに直交する項が加えられたものと解釈できる。予測性能に関しては、通常最尤推定量を用いた予測よりもアンサンブルをとった予測の方が漸近的に良い予測を与えることを示す。また、パラメトリック・ブートストラップを用いた予測とノンパラメトリック・ブートストラップを用いた予測が、漸近的に同一の予測性能をもっていることを明らかにする。パラメトリック・ブートストラップでは乱数の発生が困難な場合があるが、ノンパラメトリック・ブートストラップは、容易に実現できるので、このような場合は後者が有効である。

本論文で議論したアンサンブル学習は、繰り返しの学習を必要とするため、大きな計算リソースを要求する。計算機の発達した現代では、このような複雑な計算をしても良い予測を得る手法が現実的に必要とされる状況がある。アンサンブル学習に対する統計的な観点での理論研究は、十分に行われているとはいえず、本論文はその基礎を与えるものである。