

実データで学ぶ人工知能講座

演習課題：PLA_noisy

Matthew J. Holland*
大阪大学 産業科学研究所

問題 1. 前の PLA の例 (PLA_dissect-JA) と比較して、データの分布がどのように変わっているか（線型分離可能性など）。

問題 2. 今のデータ生成関数では、データとは独立に、正規分布に従うノイズを発生させているが、そのノイズ分布のパラメータは何か（コード内の変数名）。分散パラメータを変えて、再びデータを生成してみること。分散の大小の影響について説明すること。

問題 3. 元のデータに依存しないノイズを自分で実装してみること。たとえば、ラベルが無作為に反転する確率が $0 < \delta < 1$ （このデルタは自分で設定できるパラメータ）。

問題 4. 元のデータに依存するノイズを自分で実装してみること。たとえば、入力空間におけるデータの「中心」をその標本平均であるとする。その中心からの距離が $\epsilon > 0$ よりも大きい点のラベルが 90% の確率で反転する（このイプシロンは自分で設定できるパラメータ）。

問題 5. 各ステップにおける PLA の標本内外の識別誤差率を記録し、その推移をプロットしてみること。成績の安定性がどのように見えるか。

問題 6. PLA 安定化を図るため、スライド資料で説明している PLA の「ポケット版」を実装し、走らせてみる。結果がどのように変わるか。

問題 7. 従来の PLA と同様に、ポケット型の PLA の標本内外の識別誤差率を各更新ステップで記録し、その推移をプロットしてみること。従来とポケット型で、どのような異なるか。

*作者の連絡先：matthew-h@ar.sanken.osaka-u.ac.jp.