実データで学ぶ人工知能講座
演習課題：PLA_bounds

Matthew J. Holland*
大阪大学 産業科学研究所

「過剰期待損失の分布」に関する問題

問題 1. 期待損失 R は近似的に計算していると述べたが，どのように近似しているか説明すること．

問題 2. 過剰リスクの分布に基づいて計算した分位値から，PLA の学習能力について何が言えるか．

問題 3. また，リスク自体の分布に基づいて計算した分位値から，PLA の学習能力について何が言えるか．

問題 4. 過剰リスクの分布を実際に見ると，理論上の予想と矛盾するか．矛盾しないならば，その予想が「ぴったり」か,「甘い」か,「悲観的」か，適切な表現を選び，また選んだ理由も説明すること．

問題 5. 上記の数値実験結果に基づいて，標本内の誤り率が標本外の誤り率を少なくとも 2 ポイント上回る確率はいくらか．

問題 6. 既定値から，PLA の反復回数の上限を 2 に下げてみること．過剰リスクの分布がどのように変わるか．リスク自体の分布がどのように変わるか．

問題 7. 従来の PLA ではなく，PLA の「ポケット版」を上記の数値実験で走らせてみること．過剰リスクの分布がどのように変わるか．リスク自体の分布がどのように変わるか．

問題 8. 標本数 n を減らし，上記の数値実験を再度実行すること．過剰リスクの分布がどのように変わるか．リスク自体の分布がどのように変わるか．

「過剰リスクの裾の上界」に関する問題

問題 9. 過剰リスクの上界から，過剰リスクが $\varepsilon > 0$ を上回る確率 $\mathbf{P}\{R(\widehat{h}) - \widehat{R}(\widehat{h}) > \varepsilon\}$ の上界を導き出すこと（指数関数的に減る）．これを**過剰リスクの「裾」**の上界と呼ぶ．

問題 10. ノートブック中の関数 `err_gen_VC_perceptron` を参考にしつつ，前の質問で導出した過剰リスクの「裾」の上界を安全に計算する関数を定義し，`tail_gen_VC_perceptron` と名づけること．

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

問題 11. 上記のコードブロックでの PLA を用いた数値実験の結果 (‘perf‘の中身）に基づいて，$\mathbf{P}\{R(\widehat{h}) - \widehat{R}(\widehat{h}) > \varepsilon\}$ を ε の関数として近似的に計算すること．等間隔で $0 < \varepsilon < 0.3$ の範囲内で関数値を求めて，そのグラフを可視化すること．

問題 12. 過剰リスクの「裾」の上界も ε に依存するので，前の質問を踏まえて，同じ ε の範囲内でこの上界を計算し，そのグラフを先ほどの $\mathbf{P}\{R(\widehat{h}) - \widehat{R}(\widehat{h}) > \varepsilon\}$ の近似のグラフとともに可視化すること．理論上の上界との矛盾はあるか．理論上の上界が「タイト」が「ゆるい」か，描いたグラフに基づいて答えること．

「標本複雑度の上界」に関する問題

問題 13. パーセプトロンの過剰リスクの上界を踏まえて，過剰リスクを少なくとも $(1-\delta)$ の確率で ε 以内に収めるためには，標本数 n が最低いくらあれば十分か．ヒント：これまでに見てきた上界から，n の下界を導き出せば良く，この下界自体は n に依存する．最低必要となる標本数をパーセプトロンの**標本複雑度** (sample complexity) と呼ぶので，ここで導出する n の下界が標本複雑度の上界にあたる．

問題 14. ノートブック中の関数 `err_gen_VC_perceptron` を参考にしつつ，前の質問で導出した標本複雑度の上界を安全に計算する関数を定義し，`sample_VC_perceptron` と名づけること (n と d と δ と ε を引数として渡す必要がある)．

問題 15. 任意の $0 < \delta < 1$ と $\varepsilon > 0$ に対して，過剰リスクを確率 $(1-\delta)$ で ε 以内に抑えるのに十分な標本数 n を求めること．具体的には，先ほどの標本複雑度の上界から近似的に計算すること．これを $n(d, \varepsilon, \delta)$ と書く．ヒント：上界自体を n の関数として，その関数の不動点を反復的に求めれば良い．

問題 16. 前の質問を踏まえて，反復的に計算する $n(d, \varepsilon, \delta)$ に着目する．$\delta = 0.1$ と $\varepsilon = 0.1$ を固定して，入力次元 d を増やすことで，$n(d, \varepsilon, \delta)$ がどのように増えていくが調べてみること．結果として，パーセプトロンの場合，d_{VC} のだいたい何倍あれば n が理論的に十分といえるか．

問題 17. 理論と実践の乖離を調べる課題：

A. $d = 3$ と $\delta = 0.1$ と $\varepsilon = 0.1$ を固定し，係数 $k \in \{1, 10, 100, 1000, 10000\}$ の各々の値に対して，上記の数値実験と同様に，PLA に $n = k\, d_{\mathrm{VC}}$ だけの訓練データを与えて，標本内外の識別誤差率を記録しておくこと．

B. この成績に基づいて，各 k に対して，過剰リスクの $(1-\delta)$ の分位値を叩き出すこと．

C. 係数 k の関数として，過剰リスクの $(1-\delta)$ の分位値のグラフを可視化してみること．

D. このグラフが，n がいくらのときに ε を下回るか調べてみること．

E. この結果として，パーセプトロンの場合，d_{VC} のだいたい何倍あれば n が**実際**に十分といえるか．

F. 結論として，理論と実際の挙動とで，どの程度の乖離が見られるか．