

### レポート課題 3: $k$ -最近傍法

提出締切 2月2日(火) 18:00, 提出先: A-333

目的:  $k$ -最近傍法 (nearest neighbor rule) と交差確認法 (cross validation) を理解し, 手書き数字文字を認識するソフトウェアを作成, その性能を分析する.

以下の課題を試みる前に, 教科書 p.185–187 のコードを走らせ, 手書き文字の認識ができることを確認せよ.

レポート課題 :

1.  $k$ -最近傍法 (nearest neighbor rule), 交差確認法 (cross validation) について, それぞれの概念の意味を自分の言葉で簡潔に説明せよ (合わせて A4 用紙 1 枚程度).
2. 教科書 p.185–187 の各コードに, 分かりやすいコメントをつけよ. C や Python 等, 別の言語で同等なものを独自に作成した場合は, そのコードにコメントを付ければよい.
3. 横軸に  $k$ , 縦軸に  $\mathcal{E}_k$  (近傍数  $k$  に対する  $h$  種類の誤認識率の平均) をプロットした図を描け (点を線で結ぶこと).  $\mathcal{E}_k$  は教科書 p.184 図 13.4 の手順 2.(b) で得られる. 教科書には誤植があるので注意すること.

$$\mathcal{E}_k := \frac{1}{h} \sum_{l=1}^h \mathcal{E}_k^{(l)}$$

4. p.185 のコード p185.m では, `kk=[1:10]` と記述されており,  $k$  は 1 から 10 の値が試されている. また p.186 の `cv.m` では,  $h=5$  に設定されている.  $h$  の値を変えて, 何通りか試し, それぞれについて図を描け (3. と同じグラフ.  $h$  の値だけが異なる).  $k=10$  までとなっているので, 必要であれば, より大きな  $k$  の値も試してみよ.
5. この手法はどのくらい有効に使えるか, 考察せよ.
6. レポートの最後には感想を記述してほしい. 特に, 理解できた点, 理解できなかった点, 疑問点などを, 具体的に 箇条書きしてほしい.

注意事項:

レポートは, 1 年前の自分が読んでも, 何を調べようとしているのか (目的), 得られた結果 (図) の読み方・解釈の仕方が分かりやすく書けていればよい (コレは簡単ではない). レポートは提出期限を過ぎても 2月22日(月) 18:00 までは受け付ける が, それ以降は一切受け付けない.