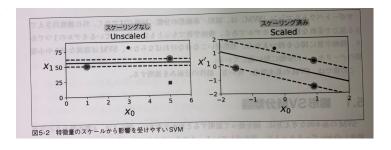
5章 サポートベクトルマシン(SVM)

- 1. 線形分割可能である複数のクラスに対し、クラスの間にできる限り太い道を通すように決 定境界を定めるような考え方である。ただし、道の端はサポートベクトルを通り、決定境 界に平行な直線(平面)である。
- 2. 決定境界として引いた直線や平面に最も近いインスタンスのこと。
- 3. 右図のように、特徴量のスケーリングが 不十分な場合、分類器はスケールが大き い方の特徴量を偏重するような決定境界 の引き方をする(図の場合はX1を偏重)。 この時、スケーリングをしている方と比較 して汎化性能が明らかに小さい。このよう に、スケーリングが重要なのは、その方 が汎化性能が高くなるためである。



- 4. 確信度も確率も出力できない ←違う。
 - SVM分類器はテストインスタンスと決定境界の距離を出力できるので、それを確信度のスコアとして使える。ただし、このスコアは直接クラスに属する推定確率に変換できない。probability=Trueを設定すると、SVMを訓練した後で、ロジスティック回帰を使って確率に較正する。この場合、SVMにpredict_prob()、predict_log_proba()メソッドが追加される。
- 5. 特徴量数よりインスタンス数の方が多いので、主問題を使うべき
- 6. yは増やすべき。また、Cは増やすべき。
- 7. p714参照。
- 8. ファイル「code_5_8.ipynb」、図1~4参照のこと。確かにどれも似たような決定境界になる。
- 9. ファイル「code_5_9.inpynb」参照のこと。プロセスの高速化の方法が思いつかず、解答をほぼ丸写し。訓練セットの前から1万インスタンスを使ってSVM分類器を訓練し、その分類器を用いてランダムサーチでパラメータ探索している。一部の訓練インスタンスで訓練を始めることが重要のようだ。しかし、このプログラムにはOneVsRestClassifier()がどこにも使われておらず、なぜOVA法だといえるのか分からなかった。
- 10. ファイル「code 5 10.ipynb」参照のこと。

