

画像認識2021 期末レポート

2110370801 澤祐里

提出日 7/18

AdaBoostとは

- アンサンブル学習の手法の一つであるブースティングの代表的な手法。

アンサンブル学習とは

- 異なるモデルの組み合わせにより、高い予測性能を持ったモデルを作成すること。

例えば、10人の専門家が予測をするとする。アンサンブル学習を利用して、10人の専門家の予測を戦略的に組み合わせると、1人1人の予測よりも正確な予測が得られる。

ブースティングとは

- 誤答から学習する弱学習器(モデル)により、強力なモデルを構築すること。

AdaBoostの特徴

- モデルを複数組み合わせているため、正確な予測結果が得られやすい。
- 外れ値(同じクラスにおいて、他の値から大きく外れた値)などに弱く、過学習(識別境界が外れ値に過剰に適合してしまい、その結果として精度が悪くなってしまうこと)になりやすい。

AdaBoostの流れ

1. 全ての訓練データが等しく重み付けされた訓練データセット D 全体で、ベース分類器(決定木など)を使って、1つ目の弱学習器 L_1 を訓練させる。
2. 弱学習器 L_1 で誤分類されたデータ D_1 の重みを大きくして、二つ目の弱学習器 L_2 を訓練する。
3. 弱学習器 L_1, L_2 で結果が異なる訓練データ D_2 を洗い出して、 D_2 の重みを大きくしてから、三つ目の弱学習器 L_3 を訓練する。
4. 弱学習器 L_1, L_2, L_3 を多数決により組み合わせる。

- 例えば、以下の図1のような○と×の2値分類の訓練データセットを考えたときに、AdaBoostを使うとどうなるかを考える。
グラフのx1,x2軸はデータがどの位置に存在するかを表す。

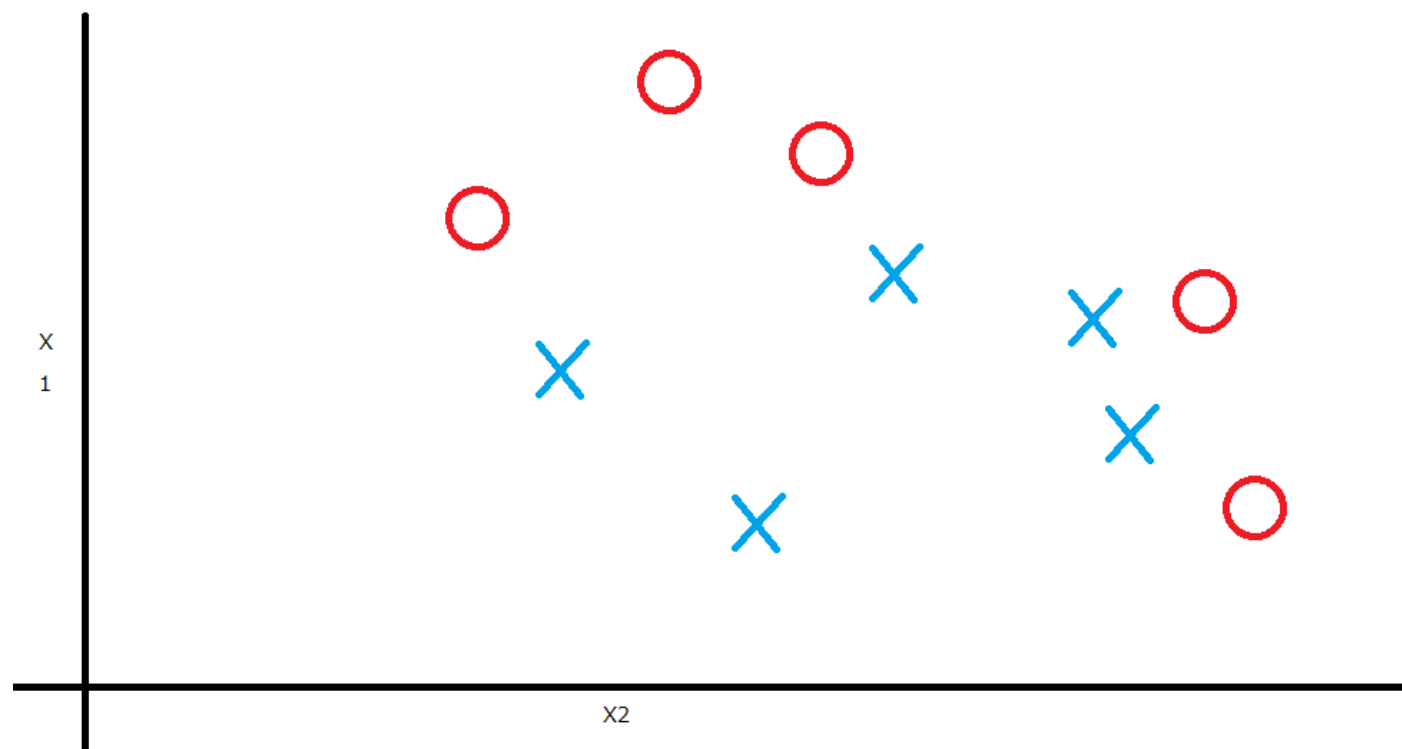


図1.2値分類の訓練データセット

- 以下の図2では、すべての訓練データが等しく重みづけされており、ベース分類器(決定木など)を使い、○と×をできるだけうまく分類しようとしている。

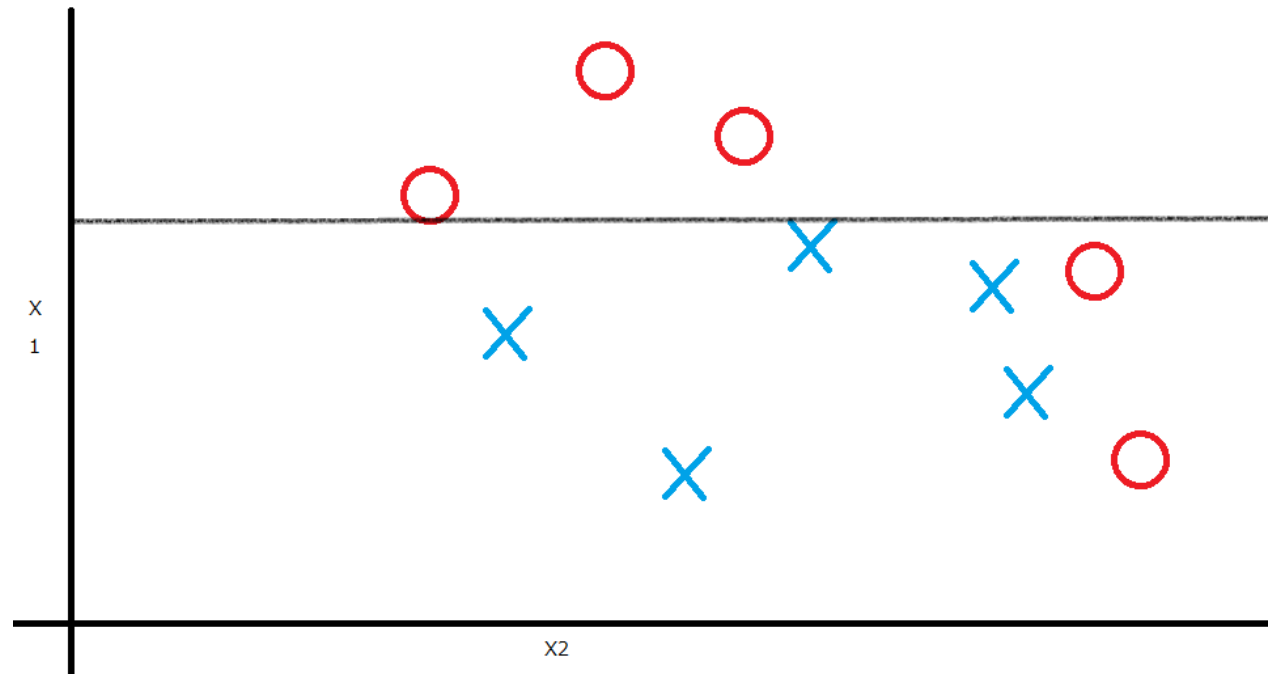


図2. 弱学習器L1の訓練

- 図3では、前回に誤分類された二つの○の重みを大きくしている。重みの合計は常に1であるため、他のデータの重みは小さくなっている。ここでは、重みが最も大きい(分類が難しい)データを重視する。

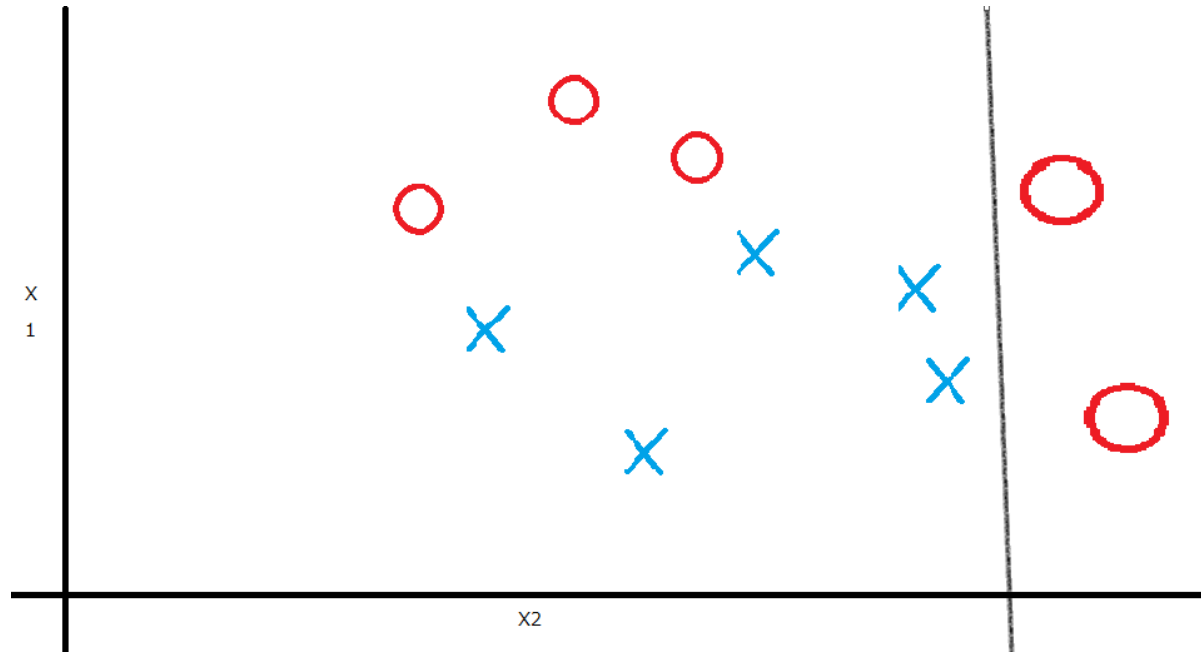


図3. 弱学習器L2の訓練

- 図4では、2回目に誤分類された3つの○の重みを大きくしている。×の重みが2回分小さくなっているため、少し×が入っているものの、○全てが入るような識別境界ができた。

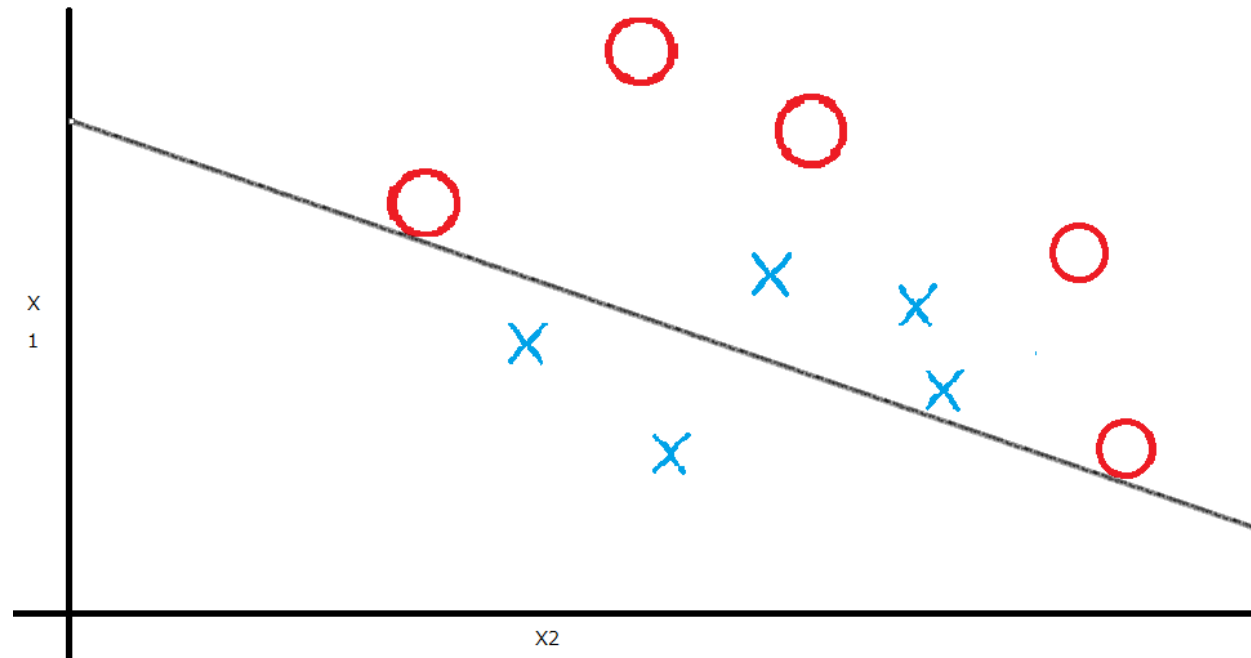


図4. 弱学習器L3の訓練

- AdaBoostアンサンブルが3回のブースティングのみで構成されているならば、図5のように、訓練された3つの弱学習器の多数決で組み合わせることになる。

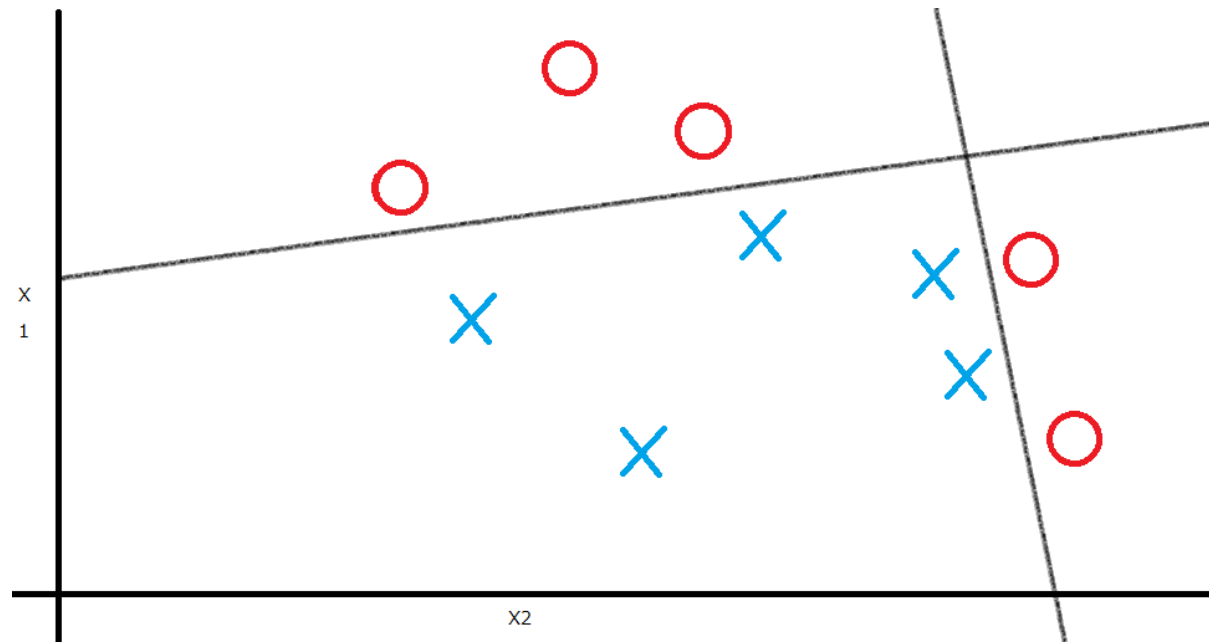


図.5 最終的な結果

AdaBoostのハイパーパラメータ

- クラスの分類手法(**ベース分類器**)・・・一般的には、決定木が使われる。

決定木を使う場合は、決定木のハイパーパラメータである、木の深さ(**深さを増やすと学習率が上がるが、学習に時間がかかる**)などを設定する必要がある。

決定木を使わない場合は、説明変数をオートスケーリング(**各変数から平均値を引いて平均を0にし、各変数を標準偏差で割って標準偏差を1にする操作**)するかどうかなどの設定をする必要がある。

- 弱学習器(**モデル**)の数・・・これまでのスライドでは、3つの弱学習器を使ってきた。

参考文献

- <http://univprof.com/archives/16-07-01-4380227.html>
- <https://qiita.com/R1ck29/items/50ba7fa5afa49e334a8f>