機械学習概論　第 2 回レポート

05-211525　齋藤駿一

2022 年 7 月 28 日

1　前処理

まず，訓練データ（dry_bean_train_data.csv）およびテストデータ（dry_bean_test_data.csv）の要素を次のように前処理した．

1. 17 個の特徴量の数値をそれぞれ，平均 0，分散 1 となるように変換した．
2. 7 つの種類を 0 から 6 までの整数値に変換した．

2　サポートベクトルマシンによる学習

まず，訓練データを用いてサポートベクトルマシン（SVM）の学習を行い，テストデータの特徴量からその種類を推測させた．その際，SVM のカーネルは線形カーネルとした．また，コストパラメータ C とカーネル幅 γ は 10^a $(a = -3, -2, -1, 0, 1, 2)$ の中から選び，訓練データを用いて 10-fold の交差検証を行い，最適と判断された量を採用した．

その結果，交差検証での正答率は 92.4%，テストデータでの正答率は 93.1% となった．

3　多層ニューラルネットによる学習

次に，SVM のかわりに多層ニューラルネットを用いて学習を行った．前処理として，訓練データとテストデータにおいて，7 つの種類を one-hot coding に直した．ここでは，以下ようにニューラルネットを構築した．まず，入力ユニット 16 個（特徴量の数と同じ）を 30 個の中間ユニットに全結合させ，その活性化関数を tahn とした．次に，その中間ユニットを 7 個（種類の数と同じ）の出力ユニットに全結合させ，その活性化関数を softmax とした．また，損失関数は交差エントロピーとし，最適化アルゴリズムは SGD とした．そして，これをバッチサイズ 10 で 300 エポックの間学習させた．

その結果，最終的な正答率は，訓練データで 94.4%，テストデータで 93.6% となった．

その後，バッチを正規化したり，層の数を増やしたり，Early stopping（訓練データを 9:1 に分

割して前者を学習に用い，後者を過学習する前に学習を止めるために用いる）をしたりすることで，精度の向上を試みた．しかし，訓練データを減らしたことによる精度の低下のため，上述の精度を超えられなかった．

また，ガウス過程を用いて学習させることも試したが，こちらは計算時間が非常に長くなってしまい，断念した．