

データセットの取得

昨日まで、python による ADALINE の実装を行って来たが、今日はその続きとして、新しいデータセットを用いて ADALINE の汎化性能を検証していきたいと思う。

そこで、まず新しいデータセットを探す。

これまで使用してきた Iris データセットは csv 形式で保存されていたので、今回探すデータセットのフォーマットも csv 形式のものに限定する。

始めに、次のサイトに掲載されていたデータセットを用いようと思う。

[scikit-learn に付いてくるデータセット 7 種類を全部まとめてみた](#)

データセットのダウンロード

サイトに掲載されていたデータセットは scikit-learn に付属しているものであるが、今後使いやすくするために、csv 形式で保存しておく。

そのために、まずダウンロードしたデータを csv 形式で保存するプログラムを組んだ。

```
def SKLearnDataSetsLoad(DataSetName):
    """scikit-learnに用意されているデータセットをロードする関数

    Param
    -----
        DataSetName (str):
            ロードするデータセット名

    Return
    -----
        df (PandasDataFrame):
    """
    if DataSetName is "boston":
        data = load_boston()

    df = pd.DataFrame(data.data, columns=data.feature_names)

    return df

def CreateDataSetCSV(df, path):
    try:
        df.to_csv(path)
    except:
        print("CSV ファイルに書き込むことができませんでした。")
```