Datamining level3,4 205713B 朝比奈太郎 2022/05/08

Level3

採用したモデル: sklearn.svm.LinearSVC

クラス定義: sklearn.svm.LinearSVC(*penalty='l2'*, *loss='squared\_hinge'*, *\**, *dual=True*, *tol=0.0001*, *C=1.0*, *multi\_class='ovr'*, *fit\_intercept=True*, *intercept\_scaling=1*, *class\_weight=None*, *verbose=0*, *random\_state=None*, *max\_iter=1000*)[[source]](https://github.com/scikit-learn/scikit-learn/blob/baf828ca1/sklearn/svm/_classes.py#L11)

ハイパーパラメータの説明

C: float型。Default = 1.0。正則化パラメータ。正則化の強さはCに反比例する。Cは必ず正の値でなければならない。

max\_iter: int型。Default = 1000。実行する反復処理の最大数。

Level4

Code

|  |
| --- |
| import pandas as pd  from sklearn.svm import LinearSVC  from sklearn.model\_selection import cross\_val\_score  from sklearn.model\_selection import KFold  df = pd.read\_csv("beer.csv")  #print(df)  X = df[["OG", "ABV", "pH", "IBU"]]  Y = df["style"]  """  print("X.shape = ", X.shape) shape -> (行数, 列数)をタプルで取得  print("Y.shape = ", Y.shape)  print(X.loc[0]) loc -> 行列と列名で位置を指定  print(Y[0])  X.shape = (150, 4)  Y.shape = (150,)  OG 11.5  ABV 4.2  pH 4.2  IBU 14.0  Name: 0, dtype: float64  Premium Lager  """  Cs = [0.5, 1.0, 1.5] #ハイパーパラメータ  k\_folds = 5 #5分割検定  for c in Cs:  model = LinearSVC(C=c)  scores = cross\_val\_score(model, X, Y, cv=KFold(n\_splits=k\_folds, shuffle=True))  #KFold(n\_splits(分割個数), shuffle(シャッフル), random\_state(乱数シード))  average = scores.mean()  print(model.fit(X, Y))  print(model.predict(X))  print(f'C = {c}: scores={scores}, average={average:.3f}') |

Codeの説明。

今回の機械学習では、Flow chartを参考にしてsklearnのLinearSVCを採択。評価にはcross\_val\_scoreを用いた。

対象となるデータセットはbeer.csvであり、数値データ(量的変数)である、"OG", "ABV", "pH", "IBU"をXとし、分類クラスに当たる”style”をYとした。ハイパーパラメータであるCを[0.5, 1.0, 1,5]でチューニングし、150のdataを5分割検定によりtraining set, testing setに分け学習を行った。scoresには0~1の値が入り、精度が高ければ高いほど、1に近づく。５分割検定したデータ(scores)の平均値をaverageという形で算出した。

分類結果と学習評価

C = 0.5のとき

分類結果

|  |
| --- |
| LinearSVC(C=0.5)  ['Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA'  'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA'  'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA'  'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA'  'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager'] |

学習評価

C = 0.5: scores=[1. 0.96666667 1. 1. 1. ], average=0.993

C = 1.0のとき

分類結果

|  |
| --- |
| LinearSVC()  ['Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA'  'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA'  'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA'  'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA'  'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'Premium Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager'] |

学習評価

C = 1.0: scores=[1. 0.96666667 0.96666667 1. 1. ], average=0.987

C = 1.5のとき

分類結果

|  |
| --- |
| LinearSVC(C=1.5)  ['Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'Premium Lager' 'Premium Lager'  'Premium Lager' 'Premium Lager' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA'  'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA'  'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA'  'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA'  'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'IPA' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager' 'Light Lager'  'Light Lager' 'Light Lager' 'Light Lager'] |

学習評価

C = 1.5: scores=[0.93333333 1. 0.96666667 1. 1. ], average=0.980

Option

実行毎にaverageが異なる値を出すことより、さらに10回試行してCが0.5,1.0, 1.5のときのaverageの平均をとってaverageの精度を上げ、Cの値が3つのうちどの場合において一番精度が高いか調べてみた。

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| C=0.5 | 0.993 | 0.987 | 0.993 | 0.987 | 0.993 | 0.987 | 0.993 | 0.993 | 0.987 | 0.987 |
| C=1.0 | 0.987 | 0,987 | 0.987 | 0.993 | 0.987 | 0.980 | 0.993 | 0.980 | 0.993 | 0.987 |
| C=1.5 | 0.980 | 0,987 | 0.987 | 0.967 | 0.987 | 0.993 | 0.993 | 0.987 | 0.993 | 1.000 |

従って、下の表より、C = 0.5のときが一番精度が高い結果となった。

|  |  |
| --- | --- |
|  | Average of aver |
| C = 0.5 | 0.990 |
| C = 1.0 | 0.9874 |
| C = 1.5 | 0.9874 |