

Python機械学習徹底演習 逆引き辞典

pandasの操作

<code>df.corr()</code>	各カラム同士の相関係数を出力
<code>df["列ラベル"].value_counts()</code>	uniqueな値の頻度を出力
<code>df.describe()</code>	各カラムの基礎統計量を出力

sklearnの操作 (model_selectionモジュール)

<code>cross_val_score(アルゴリズム,X,y,cv)</code>	交差検証するときを使う アルゴリズムは回帰, 分類どちらでも可 返り値は配列でテストへの結果
<code>GridSearchCV(アルゴリズム, パラメータ, cv,return_train_score,scoring)</code>	パラメータチューニングする際に用いる. これに対してfitを行わないと学習しないので注意 <code>return_train_score=True</code> にすると <code>train_score</code> も見られる. <code>scoring</code> を変えると回帰にも使える.
<code>GridSearchCV.fit(X,y)</code>	前に定義した辞書型のパラメータで網羅的に学習
<code>GridSearchCV.cv_results_</code>	上記の結果を辞書型で取得
<code>GridSearchCV.best_params_</code>	test scoreが最大のパラメータを取得
<code>GridSearchCV.predict(X,y)</code>	bestなパラメータでできたモデルで予測

sklearnの操作 (preprocessingモジュール=前処理)

<code>PolynomialFeatures(degree=n).fit_transform(X)</code>	n次式に変換するように変形
--	---------------

sklearnの操作 (アルゴリズム各種)

<code>ensemble.RandomForestClassifier(params)</code>	アンサンブル学習の一種で分類器のランダムフォレスト. 主なパラメータは <code>n_estimators,max_depth</code> .
<code>linear_model.Lasso(params)</code>	線形回帰モデルの一種であるラッソ回帰. 主なパラメータは <code>normalize,max_iter,alpha</code> .

sklearnの操作 (metricsモジュール=評価)

<code>accuracy_score(prediction,true)</code>	予測したクラスと正解のクラスがどのくらい合っているかを出力する
<code>mean_squared_error(prediction,true)</code>	予測した値と正解の値との平均二乗誤差を出力する