Difference between fit,fit_transform,transform and predict

Most scikit-learn objects are either transformers or models.

Transformers

Transformers are for pre-processing before modeling. Standard Scaler,MinMax Sclaler,PCA,Imputer class (like SimpleImputer for filling in missing values) and FeatureSelection classes in sklearn are an example of some transformers.

Models

Models are used to make predictions like Linear Regression model, Decision Tree model, Random Forest model,Adaboost etc. You will usually pre-process your data (with transformers) before putting it in a model.

For Transformers:

These methods are used to center/feature scale the given data. It basically helps to normalize the data within a particular range

For this, we use Z-score method.

$Z = (x - \mu)/\sigma$

We do this on the training set of data.

1.Fit(): Method calculates the parameters μ and σ and saves them as internal objects.

2.Transform(): Method using these calculated parameters apply the transformation to a particular dataset.

3.Fit_transform(): joins the fit() and transform() method for transformation of dataset.

- fit() => transform() or fit_transform()

used for Scalers, and NLP Vectorizers'

For Models:

1.fit() - It calculates the parameters/weights on training data (e.g. parameters returned by coef() in case of Linear Regression) and saves them as an internal objects state.

2.predict() - Use the above calculated weights on test data to make the predictions

- fit() => predict()

is almost used for all classifiers in SKLearn (Knn, SVC, Logistic Reg, NaiveBayes … etc)

In [ ]: