FINAL MACHINE LEARNING REPORT

Contents

NOTE ON DATA ACQUISITION

Are your features informative, discriminating, independent, explainable? Look for correlations or combinations.

4. Prepare your data to better expose the underlying data patterns to ML algorithms. Does your data need cleaning? Do outliers need to be detected and removed? Do your features need to be transformed or scaled?

WHAT IS WRONG HERE?

```python
from sklearn.metrics import mean_squared_error
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import StandardScaler
from sklearn.datasets import make_regression
from sklearn.model_selection import train_test_split

X, y = make_regression(n_features=1, noise=1)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.4)

scaler = StandardScaler()
X_train_transformed = scaler.fit_transform(X_train)

model = LinearRegression().fit(X_train_transformed, y_train)
mean_squared_error(y_test, model.predict(X_test))
```

USE PIPELINES!

```
from sklearn.pipeline import make_pipeline

model = make_pipeline(StandardScaler(), LinearRegression())
model.fit(X_train, y_train)


mean_squared_error(y_test, model.predict(X_test))
```

10. Common pitfalls and recommended practices — scikit-learn 1.0 documentation (https://scikit-learn.org/stable/common_pitfalls.html)

MODEL PERSISTENCE

- https://scikit-learn.org/stable/modules/model_persistence.html

- Simply use pickle an try it out