

QUIZ TIME

- Individual, multiple-choice questions
- Online: <http://www.socrative.com> room **1PTGB6PY**
- Open book quiz, so books and slides can be consulted
- **HAN student number**, so NOT your name, nickname or anything else.
- Quiz starts exactly at class hour and takes 10 minutes.
- Be on time and have your equipment prepared.

QUIZ RESITS

Remember: last score counts!

October 19th in R29 2.12

- 09:45 Qz 1ML basics (resit)
- 10:00 Qz 2 Data in ML (resit)
- 10:15 Qz3 Performance (resit)

You only need to come by for the quiz you want to redo

FINAL MACHINE LEARNING REPORT

Contents

1	Introduction.....	3
2	Problem statement.....	4
3	Data acquisition and exploration.....	5
4	Feedback FROM another team on your preliminary report.....	6
5	Feedback TO another team on THEIR preliminary report	7
6	Model selection, training and validation	8
7	Deploy and test.....	9
8	Conclusion	10
9	References	11
	Code appendices	12

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?

```
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 and try it out