Learning pattern through optimization
Parameters of a model are iteratively updated with respect to loss J of the training set

Hyperparameter configuration
Regularization

training set
validation set
test set

Learning pattern through optimization

Hyperparameter configuration
Model is tested on validation data in order to be able to configure hyperparameters and see the generalization performance of the model

Regularization

training set

validation set

test set

Learning pattern through optimization
Hyperparameter configuration

training set

validation set

Verifying generalization

Through adjusting the hyperparameters based on the validation loss, this data is learned implicitly. To verify your model performance it is finally run on the test set

test set

Dataset split

Less training data
.. higher variance in parameter estimates

Less validation and test data
..higher variance in performance estimate

Highly dependent on dataset and task

training set

validation set

test set

Dataset split

Less training data
.. higher variance in parameter estimates

Less validation and test data
..higher variance in performance estimate

Cross validate if possible.

80% training set

80% 20% validation set

20% test set

Sampling order

Random sampling

When splitting your dataset, do random sampling to break collection biases.

E.g. Time dependencies, sensor dependencies

Random order

Dataset with high variance

https://pixabay.com/

https://pixabay.com/

tank

Random sampling

Dataset with high variance

Your model can only depict the data you collected.

A high data variance enables generalization and makes your model less prone to data biases.

no tank

https://pixabay.com/

https://pixabay.com/

tank
sunny

Random sampling

Dataset with high variance

Your model can only depict the data you collected.

A high data variance enables generalization and makes your model less prone to data biases.

overcast
no tank

https://pixabay.com/