EMBEDDED VISION DESIGN 3

ML PERFORMANCE

HANDS ON

JEROEN VEEN

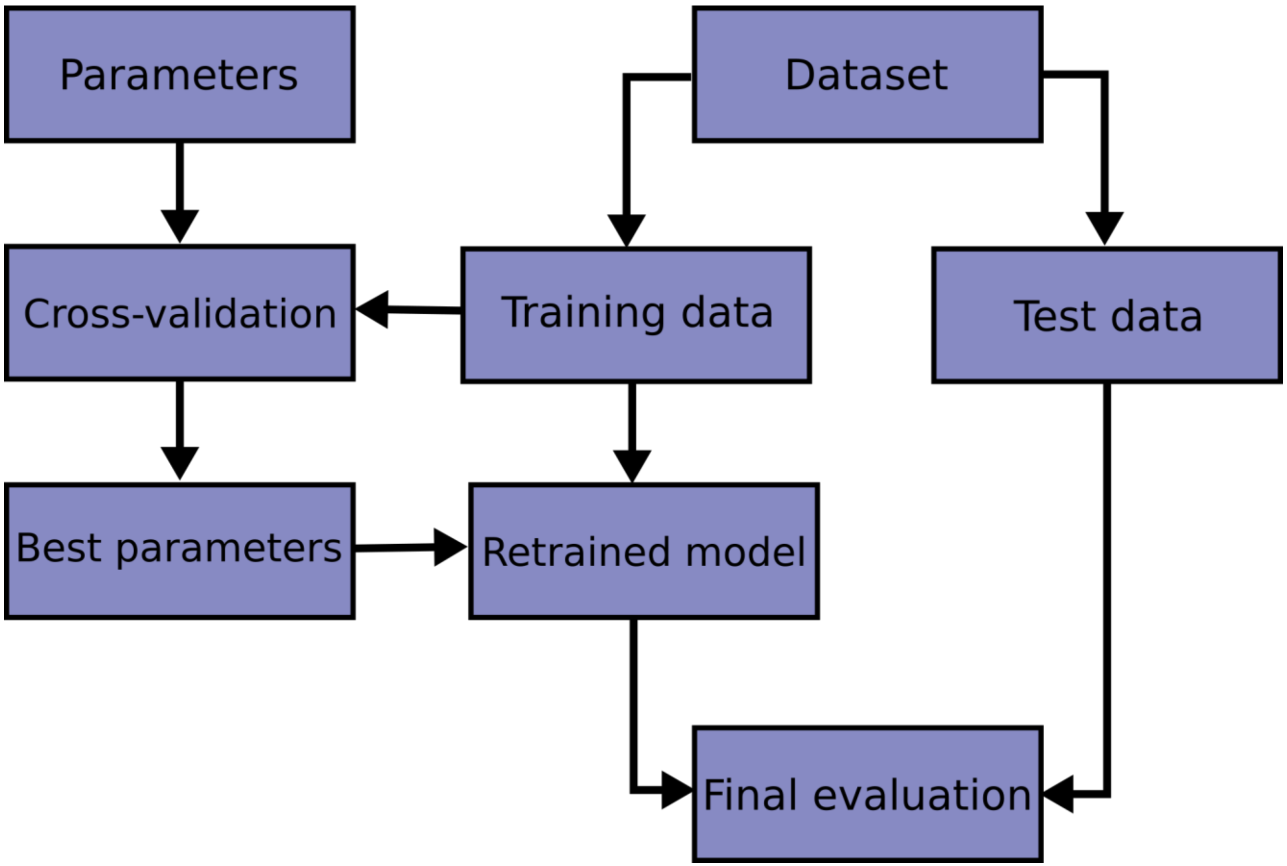
HAN_UNIVERSITY
OF APPLIED SCIENCES

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.

EVALUATING ESTIMATOR PERFORMANCE



Source: https://scikit-learn.org/stable/modules/cross_validation.html

CONFUSION MATRIX

ACTUAL

(Type I error)

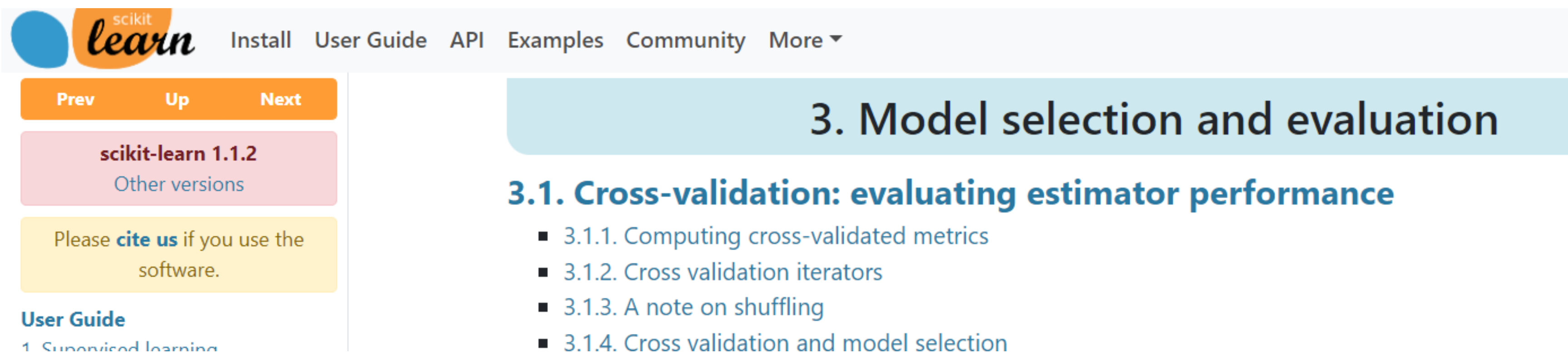
PREDICTED

True Positive (TP) Reality: A wolf threatened. Shepherd said: "Wolf." Outcome: Shepherd is a hero.	**False Positive (FP)** Reality: No wolf threatened. Shepherd said: "Wolf." Outcome: Villagers are angry at shepherd for waking them up.
False Negative (FN) Reality: A wolf threatened. Shepherd said: "No wolf." Outcome: The wolf ate all the sheep.	**True Negative (TN)** Reality: No wolf threatened. Shepherd said: "No wolf." Outcome: Everyone is fine.

Type II error)

LOOK INTO THE SCKIT USER GUIDE

Find the descriptions of the model evaluation functionality



https://scikit-learn.org/

MULTICLASS AND MULTILABEL

- extending a binary metric to multiclass or multilabel problems, the data is treated as a collection of binary problems
- several ways to average binary metric calculations across the set of classes

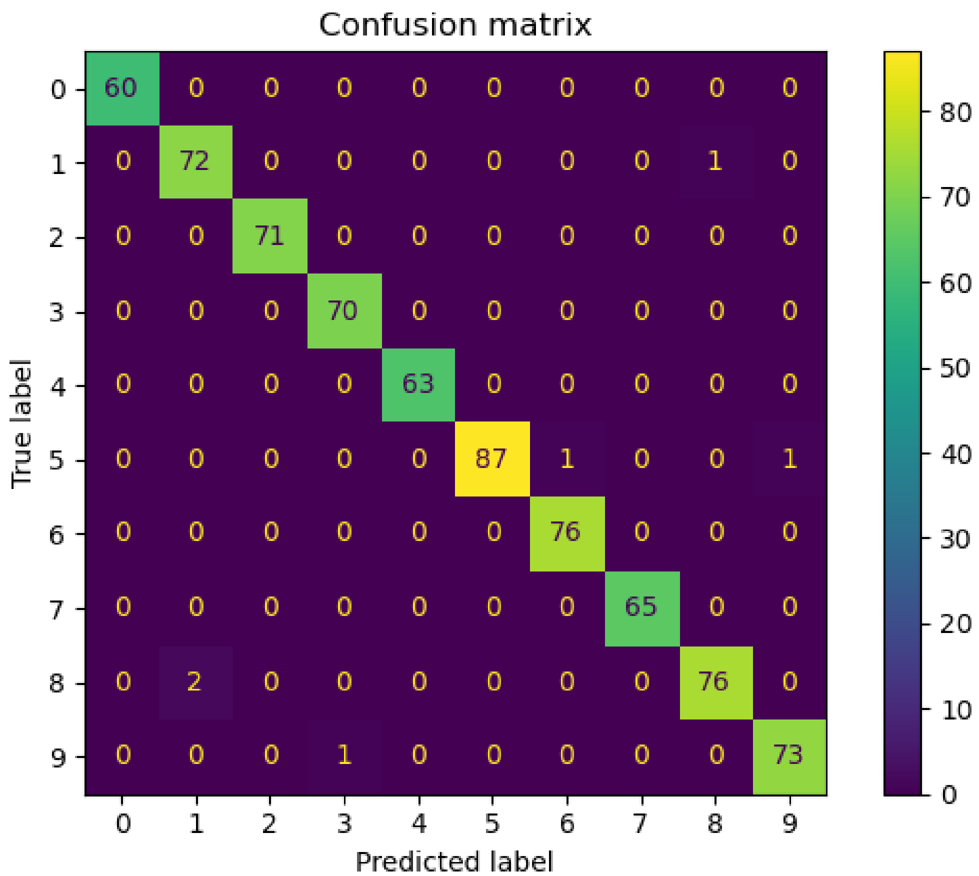
- `"macro"` simply calculates the mean of the binary metrics, giving equal weight to each class. In problems where infrequent classes are nonetheless important, macro-averaging may be a means of highlighting their performance. On the other hand, the assumption that all classes are equally important is often untrue, such that macro-averaging will over-emphasize the typically low performance on an infrequent class.
- `"weighted"` accounts for class imbalance by computing the average of binary metrics in which each class's score is weighted by its presence in the true data sample.

EXAMPLE

```python
# load a dataset, see https://scikit-learn.org/stable/datasets/index.html#datasets
digits = datasets.load_digits()

# Sample a training set while holding out 40% of the data for testing (evaluating) our classifier
X_train, X_test, y_train, y_test = train_test_split(
    digits.data, digits.target, test_size=0.4, random_state=0)

# instantiate a classifier estimator
clf = Pipeline([
    ("scaler", StandardScaler()),
    ("svm_clf", SVC(kernel="poly", degree=3, coef0=1, C=5))
    ])

# fit the classifier
clf.fit(X_train, y_train)

# show the result of the scoring method of the classifier. For SVC, the score is mean accuracy
print("mean accuracy on the given test data and labels: {}".format(clf.score(X_test, y_test)))

# compute some other metric to evaluate the model
scoring_metric = "precision_macro"
nr_of_folds = 5
score = cross_val_score(clf, X_train, y_train, cv=nr_of_folds, scoring=scoring_metric)
print("{}-fold cross validation metric {} : {}".format(nr_of_folds, scoring_metric, score))
```

Performance_01.py

CONFUSION MATRIX



Performance_02.py

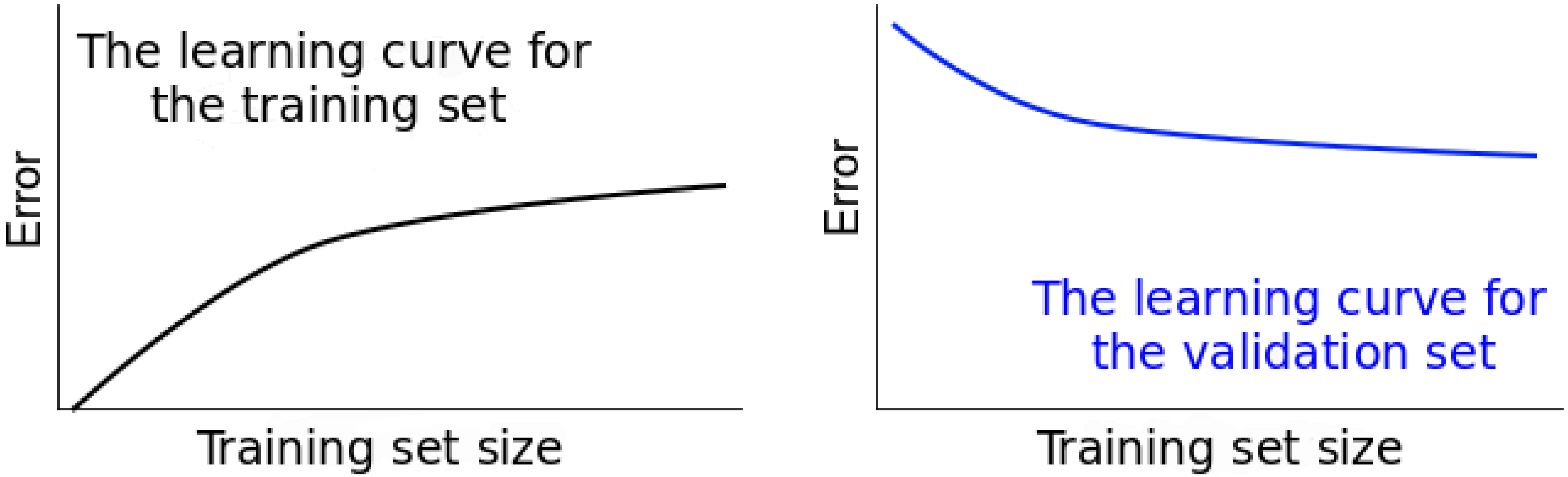
HAN_UNIVERSITY OF APPLIED SCIENCES

TRY IT FOR YOURSELF

- performance_01.py
  performance_02.py

LEARNING CURVES

- Cost as a function of the training set size (or the training iteration)
- Examine evolution of train and validation learning curves



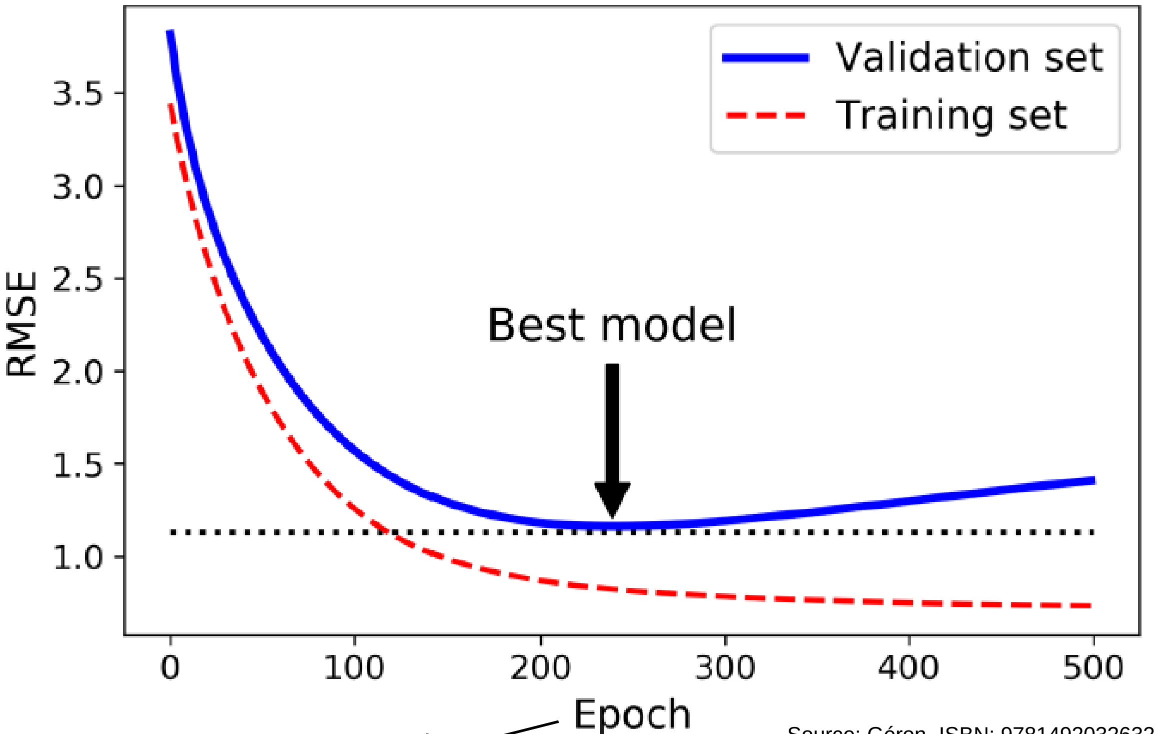
Source: https://www.dataquest.io/blog/learning-curves-machine-learning/

TRY IT FOR YOURSELF

- performance_03.py

EARLY STOPPING

- Interpretation of learning curves



Source: Géron, ISBN: 9781492032632

epoch means that each sample in the training dataset has been used to update the internal model parameters

LEANING CURVES IN REGRESSION

- Exercise 02: Plot the learning curves for polynomial regression and experiment with various degrees

- Can you interpret the curves?

- Build on Regression_01.py
  and see Géron, page 130-134