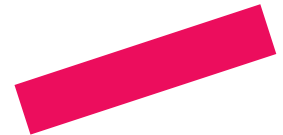


EVML3

# ML PERFORMANCE

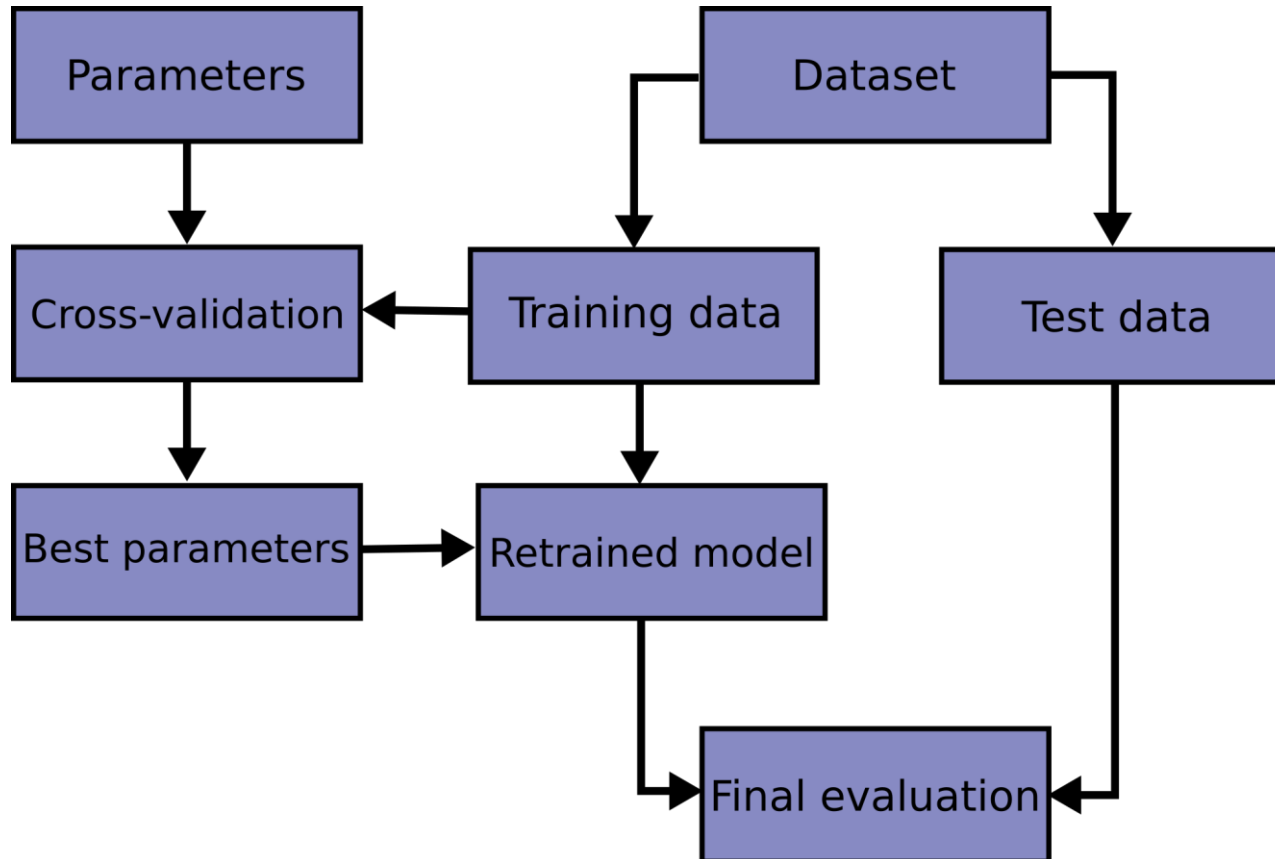
HANDS ON

JEROEN VEEN



**HAN\_**UNIVERSITY  
OF APPLIED SCIENCES

# EVALUATING ESTIMATOR PERFORMANCE



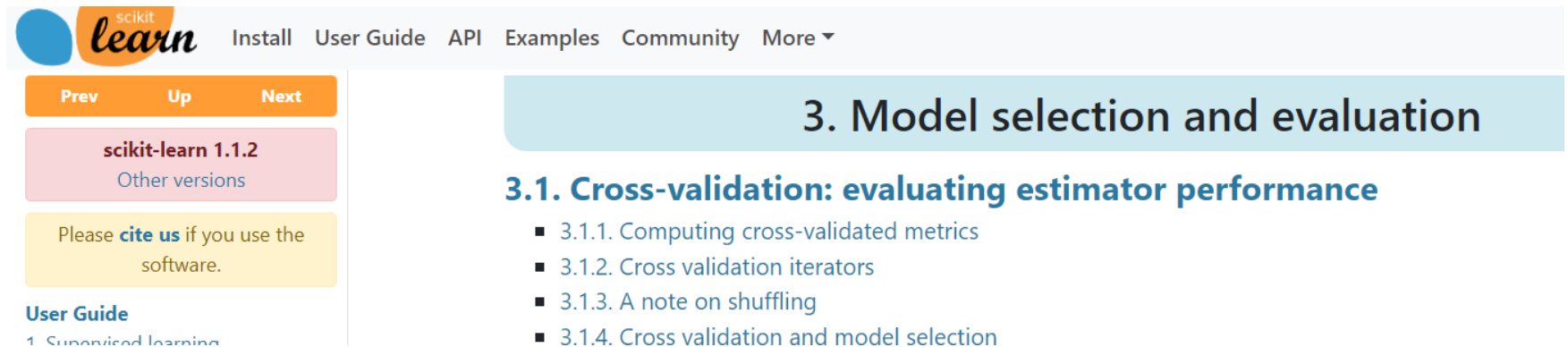
Source: [https://scikit-learn.org/stable/modules/cross\\_validation.html](https://scikit-learn.org/stable/modules/cross_validation.html)

# CONFUSION MATRIX

		ACTUAL	
PREDICTED			(Type I error)
		<b>True Positive (TP)</b> Reality: A wolf threatened. Shepherd said: "Wolf." Outcome: Shepherd is a hero.	<b>False Positive (FP)</b> Reality: No wolf threatened. Shepherd said: "Wolf." Outcome: Villagers are angry at shepherd for waking them up.
		<b>False Negative (FN)</b> Reality: A wolf threatened. Shepherd said: "No wolf." Outcome: The wolf ate all the sheep.	<b>True Negative (TN)</b> 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



The screenshot shows the scikit-learn website's navigation bar with links for Install, User Guide, API, Examples, Community, and More. The left sidebar contains navigation links for Prev, Up, and Next, the current version (scikit-learn 1.1.2), and a citation notice. The main content area displays the title '3. Model selection and evaluation' and the sub-section '3.1. Cross-validation: evaluating estimator performance', which includes a list of topics: 3.1.1. Computing cross-validated metrics, 3.1.2. Cross validation iterators, 3.1.3. A note on shuffling, and 3.1.4. Cross validation and model selection.

scikit-learn

Install User Guide API Examples Community More ▾

Prev Up Next

**scikit-learn 1.1.2**  
Other versions

Please **cite us** if you use the software.

**User Guide**  
1. Supervised learning

## 3. Model selection and evaluation

### 3.1. Cross-validation: evaluating estimator performance

- 3.1.1. Computing cross-validated metrics
- 3.1.2. Cross validation iterators
- 3.1.3. A note on shuffling
- 3.1.4. Cross validation and model selection

<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

```
# 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)

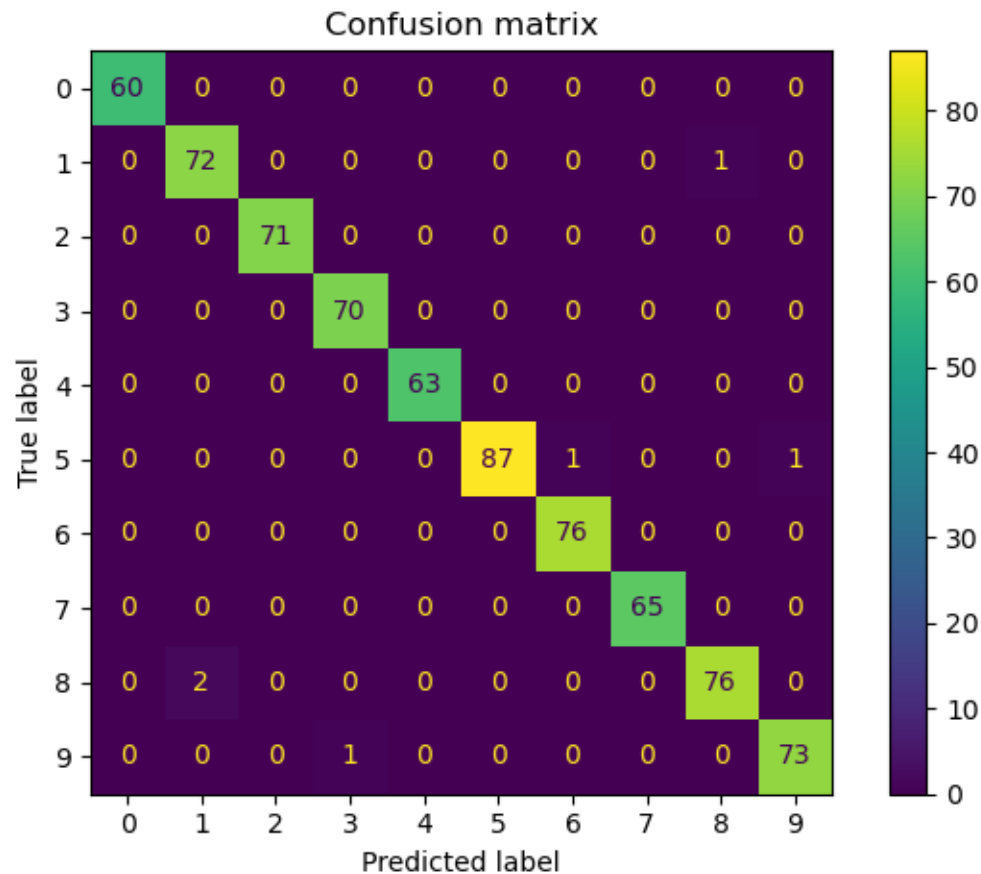
# instantiate a classifier estimator
clf = Pipeline([
    ("scaler", StandardScaler()),
    ("svm_clf", SVC(kernel="poly", degree=3, coef0=1, C=5))
])

# 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))
```

# CONFUSION MATRIX



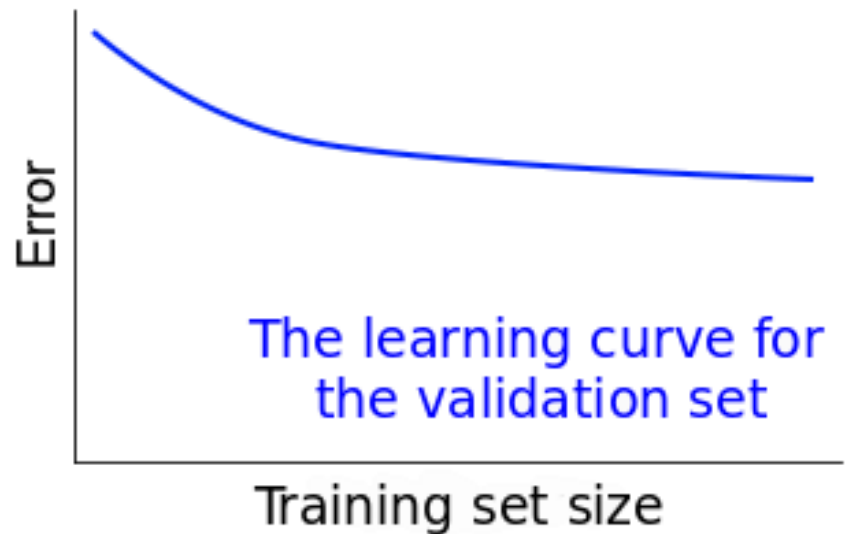
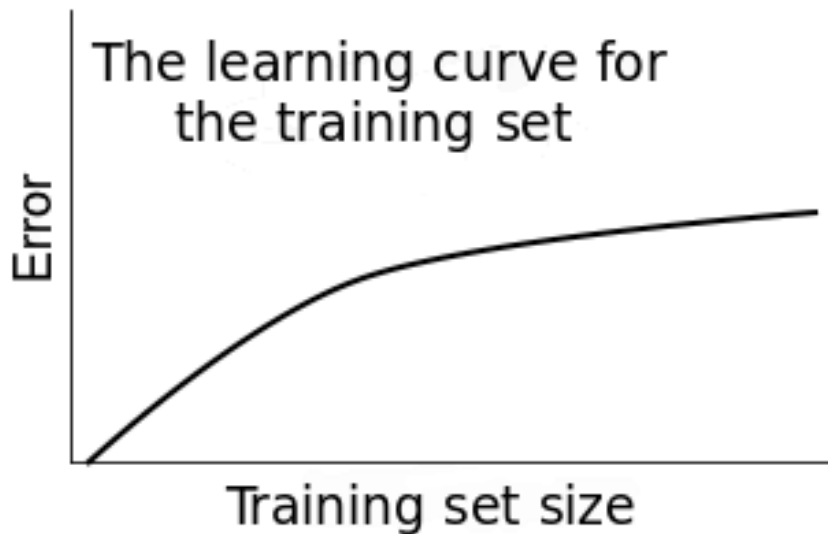
# TRY IT FOR YOURSELF

- 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



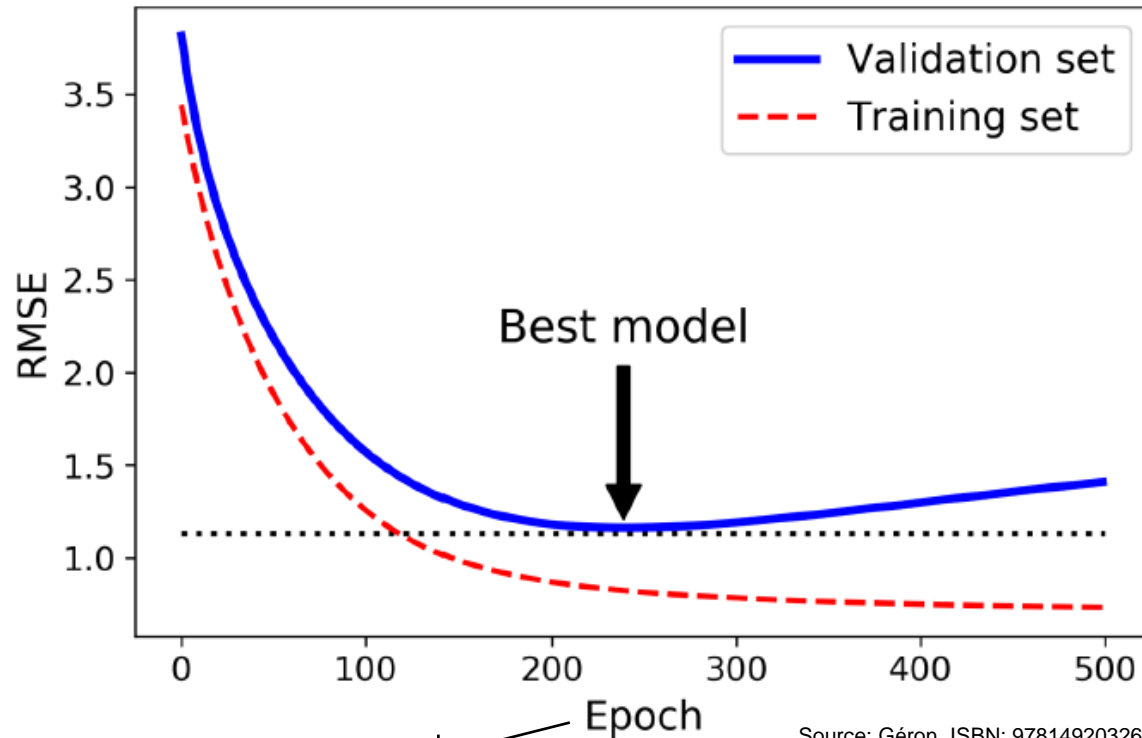
Source: <https://www.dataquest.io/blog/learning-curves-machine-learning/>

# TRY IT FOR YOURSELF

- performance\_03.py

# EARLY STOPPING

- Interpretation of learning curves



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

Epoch

Source: Géron, ISBN: 9781492032632