**Hands\_on\_machine learning**

**Chapter1 The Machine Learning Landscape**

machine learning type: training supervision: supervised, unsupervised, self-supervised, semi-supervised, reinforcement learning

1. Supervised: label(classification tasks) or target(regression tasks) is given.
2. Unsupervised: no label, clustering (novelty detection and anomaly detection), dimension reduction
3. Semi-supervised: partially labeled
4. Self-supervised learning: make use of unlabeled dataset to do classification/regression tasks.
5. Reinforcement learning/ offline learning

Batch or online learning:

1. Batch learning: offline learning
2. Online learning: incremental learning

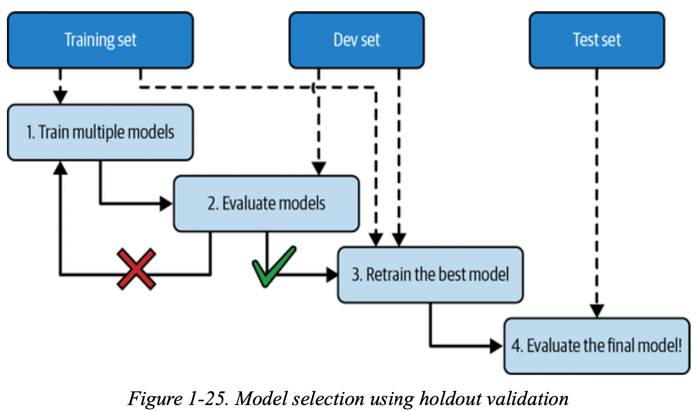
Instance based versus model-based learning:

1. Instance-based learning: no model, just instance
2. Model-based learning: model is created

Challenge of machine learning:

1. Representative sample data(sample noise when too small and potential sample bias when too large)
2. Poor-quality data
3. Model being overfitted (need regularization)

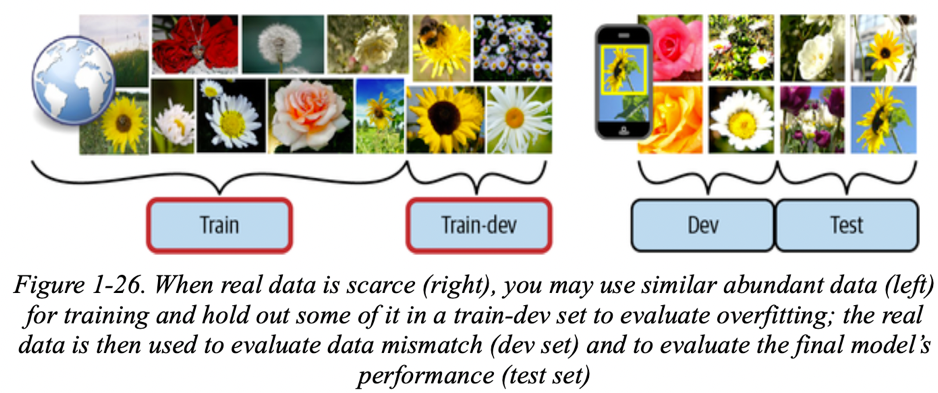
Training set, validating set and testing set

****

**Validation set**: try to solve the problem on overfitting the test set

**Cross-validation**: using many different validation sets for each model and average out the evaluations of model on validation to identify the best model to train on fully training set(to solve the problem of too few training set if the validation set is too large)

**Data mismatch**: training data is in large quantity but not representative of the data in production; solution: train-dev set



Train-dev: solve the issue of overfitting on training data

Dev: solve the issue of data mismatch.

Test set: as before, solve the issue of overfitting on dev set.