EVALUATING METHODS IN SUPERVISED LEARNING

- In order to use machine learning supervised methods in real life application, it is important to first test their performance on how well they meet the expectation of correct prediction.
- The biggest challenge here is that how can we test the behaviour of supervised methods in predicting the events which have not happened yet. The reason is that what is going to happen tomorrow is an unknown event not only for Machine learning supervised models but, also for humans.
- To address this challenge, machine learning offers very smart strategies to evaluate/test the performance of the models on forecasting future observations.
  1. Hold-out
  2. Cross-Validation
  3. Bootstrap

HOLD-OUT I

- The idea of this method is to divide the historical data set into two parts namely, training and test set. The training and test set environment under Hold-out method is illustrated in Figure 40

Figure 40: Holdout Method

The key characteristics of Hold-out method are:

1. The division of data records in training and test set is random.
2. Training set is usually 70%-80% of the data whereas, test set is formed with 20%-30% of the remaining data set.
3. Training set is used to train the model. With this data set, we let model learn. Training set is always a supervised data, *i.e.*, class labels are provided to the model for learning.
4. Test set is used for evaluating/testing the performance of the model. Test set is always unsupervised data, *i.e.*, class labels are not provided or hidden from to the model.
5. No test case is common between training set and test set.

HOLD-OUT II

Example Illustration

- Below is a su… taken from here to understand Hold-out method. Iris is a kind of flower characterised by four features namely, sepal width, sepal length, petal length and petal width.
- There are three species/classes of Iris flower: setosa, versicolor and virginica. Figure 41 illustrates the process of Hold-out method. Where, 70% of data is taken to train the model whereas, rest of the data is used for evaluation. Important is to notice that, class labels

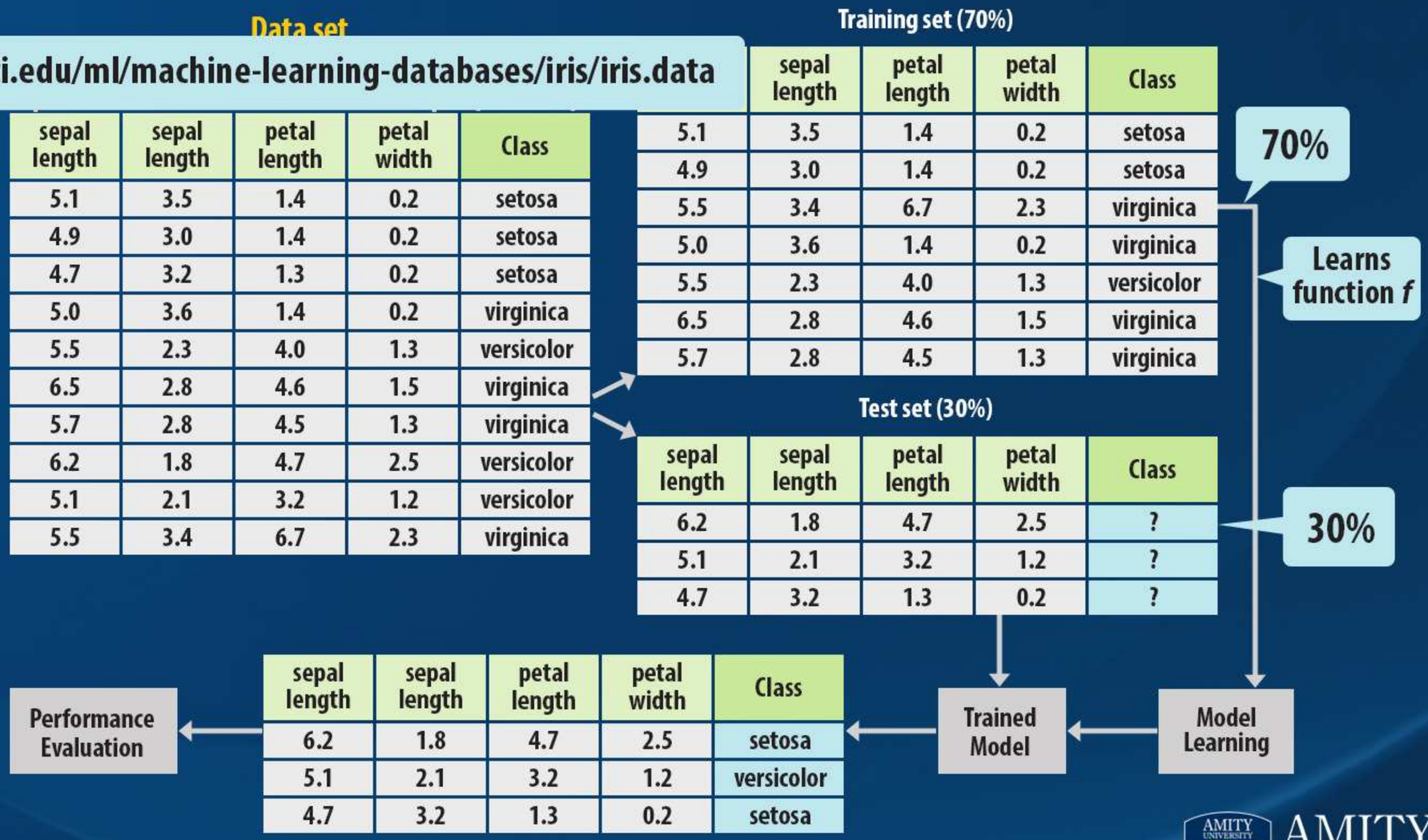


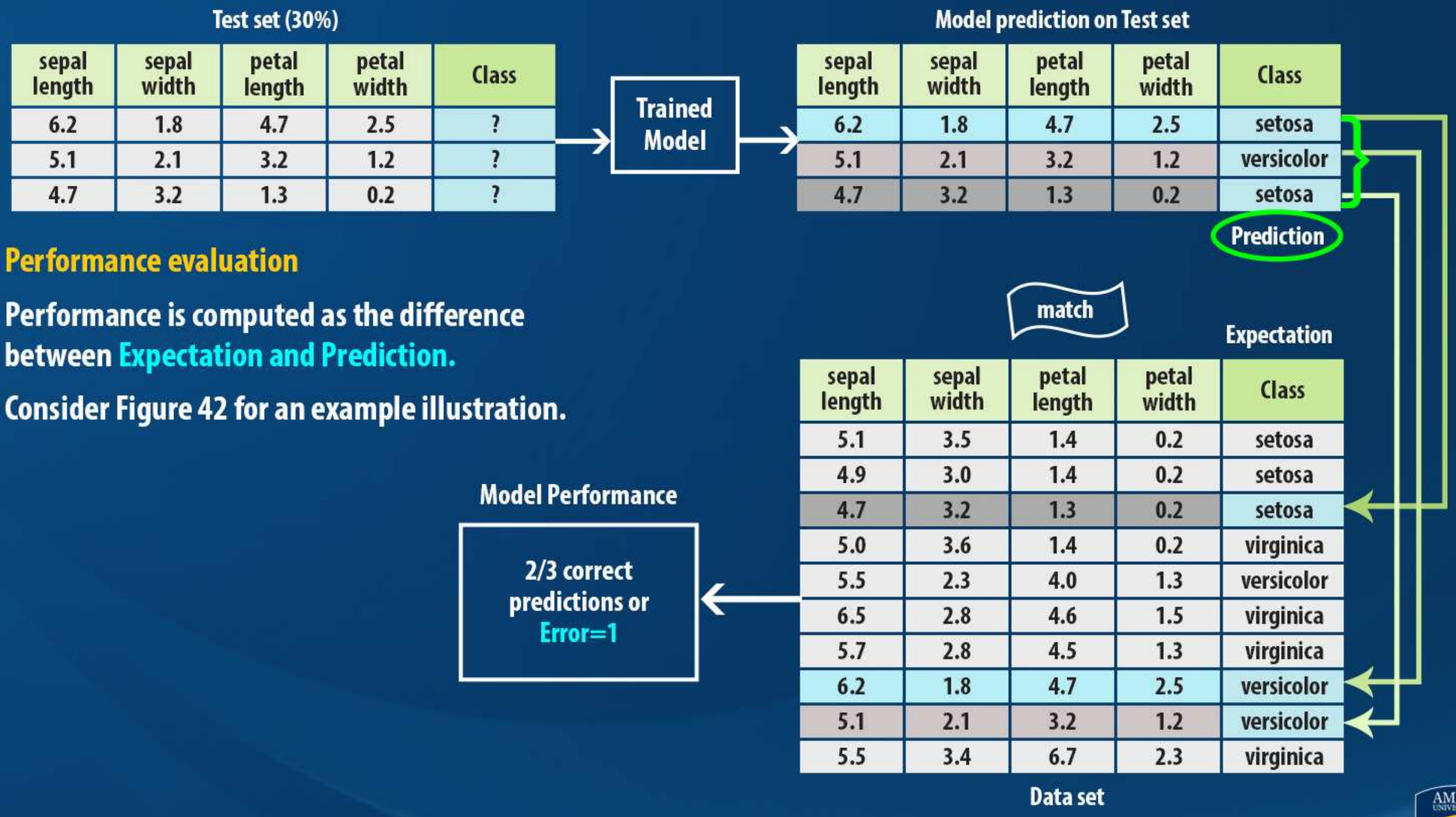
Figure 41: Holdout Method on Iris data set

Error is difference between actual and expected result.

- Lowest Error ↔ Good Performance
- Higher Error ↔ Poor Performance

Division of data set → Training of the model → Application of test set → Prediction by the model

HOLD-OUT III



Performance evaluation

Performance is computed as the difference between Expectation and Prediction.

Consider Figure 42 for an example illustration.

Figure 42: Method of performance evaluation on predictive results by the model

HOLD-OUT IV

- **Strengths and Weakness of Hold-Out**
  - **Strengths**
    1. Simplest
  - **Weakness**
    1. Performance is highly dependent upon how training and test set are created. It is difficult to authenticate the performance of the model when high proportion of data is taken for training. Whereas, low training samples may lead to poor performance and hence poor results.
    2. This method is not suitable for small data sets for the reason that we will not get enough data samples for training and test.

The above mentioned weakness are overcome by using Cross-validation method.