Machine Learning: Train Classification Models in Classification Learner App

In machine learning, choosing the right classification model is often a matter of trial and error. The Classification Learner app in Statistics and Machine Learning Toolbox™ makes it easy.

What Is Supervised Machine Learning?

Using the Classification Learner app, you can explore supervised machine learning using various classifiers. You can explore your data, select features, specify cross-validation schemes, train models, and assess results. You can perform automated training to search for the best classification model type, including decision trees, discriminant analysis, support vector machines, logistic regression, nearest neighbors, and ensemble classification.

Perform supervised machine learning by supplying a known set of input data (observations or examples) and known responses to the data (i.e., labels or classes). Use the data to train a model that generates predictions for the response to new data. To use the model with new data, or to learn about programmatic classification, you can export the model to the workspace or generate MATLAB® code to recreate the trained model.

1. Known Data \rightarrow \text{Model} \\
   Known Responses \rightarrow \text{Model}

2. Model \rightarrow \text{Predicted Responses} \\
   New Data \rightarrow \text{Predicted Responses}

Get started by training a selection of model types.
Automated Classifier Training

You can use Classification Learner to automatically train a selection of different classification models on your data.

- Get started by automatically training multiple models at once. You can quickly try a selection of models, then explore promising models interactively.
- If you already know what classifier type you want, train individual classifiers instead. See the Manual Classifier Training section.

1. On the Apps tab, in the Math, Statistics, and Optimization group, click Classification Learner.
2. Click New Session and select data from the workspace or from a file. Specify a response variable and variables to use as predictors.
3. On the Classification Learner tab, in the Classifier section, click All Quick-To-Train. This option will train all the model presets available for your data set that are fast to fit.
4. Click Train.

A selection of model types appears in the History list. When they finish training, the best percentage Accuracy score is highlighted in a box.
5. Click models in the history list to explore results in the plots.

For next steps, see the Manual Classifier Training or Compare and Improve Classification Models section.

6. To try all the classifier model presets available for your data set, click All, then click Train.

**Manual Classifier Training**

If you want to explore individual model types, or if you already know what classifier type you want, you can train classifiers one at a time, or a train a group of the same type.

1. Choose a classifier. On the Classification Learner tab, in the Classifier section, click a classifier type. To see all available classifier options, click the arrow on the far right of the Classifier section to expand the list of classifiers. The options in the Classifier gallery are preset starting points with different settings, suitable for a range of different classification problems.
To read a description of each classifier, switch to the details view.

2. After selecting a classifier, click **Train**.

Repeat to try different classifiers.

*Tip* Try decision trees and discriminants first. If the models are not accurate enough at predicting the response, try other classifiers with higher flexibility. To avoid overfitting, look for a model of lower flexibility that provides sufficient accuracy.

3. If you decide you want to try all model types or train a group of the same type, try one of the **All** options in the classifier gallery.

**Compare and Improve Classification Models**
1. Click models in the history list to explore the results in the plots. Compare model performance by inspecting results in the scatter plot, confusion matrix, and ROC curve. Examine the percentage accuracy reported in the history list for each model.

2. Select the best model in the history list and then try including and excluding different features in the model. Click Feature Selection.

    ![Feature Selection](image)

Try the parallel coordinates plot to help you identify features to remove. See if you can improve the model by removing features with low predictive power. Specify predictors to include in the model, and train new models using the new options. Compare results among the models in the history list.

    ![PCA](image)

You can also try transforming features with PCA to reduce dimensionality.

3. To improve the model further, you can try changing classifier parameter settings in the Advanced dialog box, and then train using the new options. To learn how to control model flexibility, see Choose Classifier Options.

4. If feature selection, PCA, or new parameter settings improve your model, try training All model types with the new settings. See if another model type does better with the new settings.
Tip To avoid overfitting, look for a model of lower flexibility that provides sufficient accuracy. For example, look for simple models such as decision trees and discriminants that are fast and easy to interpret. If the models are not accurate enough at predicting the response, choose other classifiers with higher flexibility, such as ensembles.

The figure shows the app with a history list containing various classifier types.