**WEKA EXERCISES**

**Make Predictions in Weka with the diabetes dataset**

After you have found a well performing machine learning model and tuned it, you must finalise your model so that you can make predictions on new data.

In this exercise you will discover how to finalise your machine learning model, save it to file and load it later in order to make predictions on new data.

After this exercise you will know:

* How to train a final version of your machine learning model in Weka.
* How to save your finalised model to file.
* How to load your finalised model later and use it to make predictions on new data.

This tutorial is broken down into 4 parts:

* Finalize Model where you will discover how to train a finalized version of your model.
* Save Model where you will discover how to save a model to file.
* Load Model where you will discover how to load a model from file.
* Make Predictions where you will discover how to make predictions for new data.
* The tutorial provides a template that you can use to finalise your own machine learning algorithms on your data problems.

We are going to use the Pima Indians Onset of Diabetes dataset (diabetes.arff file from WEKA dataset folder, I have also added this to SOL). Each instance represents medical details for one patient and the task is to predict whether the patient will have an onset of diabetes within the next five years. There are 8 numerical input variables, and all have varying scales.

We are going to finalize a logistic regression model on this dataset, both because it is a simple algorithm that is well understood and because it does very well on this problem.

## 1. Finalize a Machine Learning Model

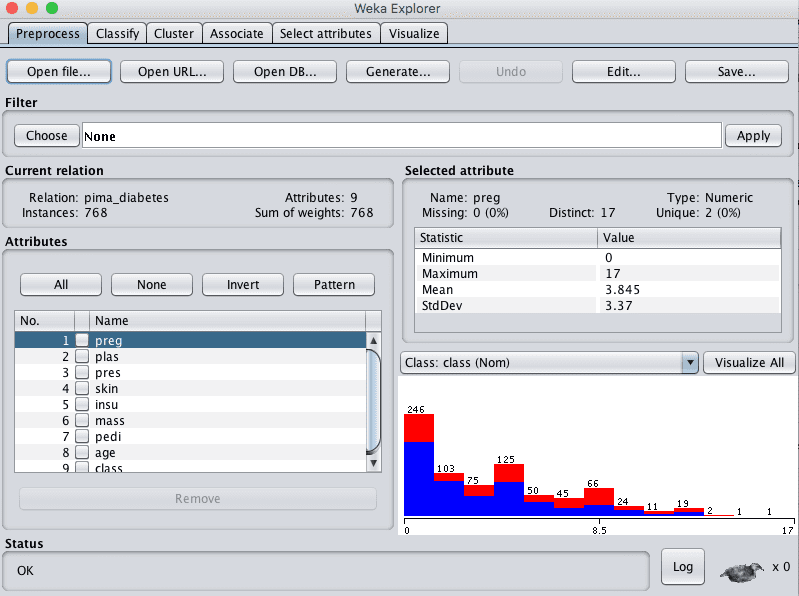
Once you have gone through all of the effort to prepare your data, compare algorithms and tune them on your problem, you actually need to create the final model that you intend to use to make new predictions.

Finalising a model involves training the model on the entire training dataset that you have available.

1. Open the Weka GUI Chooser.

2. Click the “Explorer” button to open the Weka Explorer interface.

3. Load the Pima Indians onset of diabetes dataset from the data/diabetes.arff file.



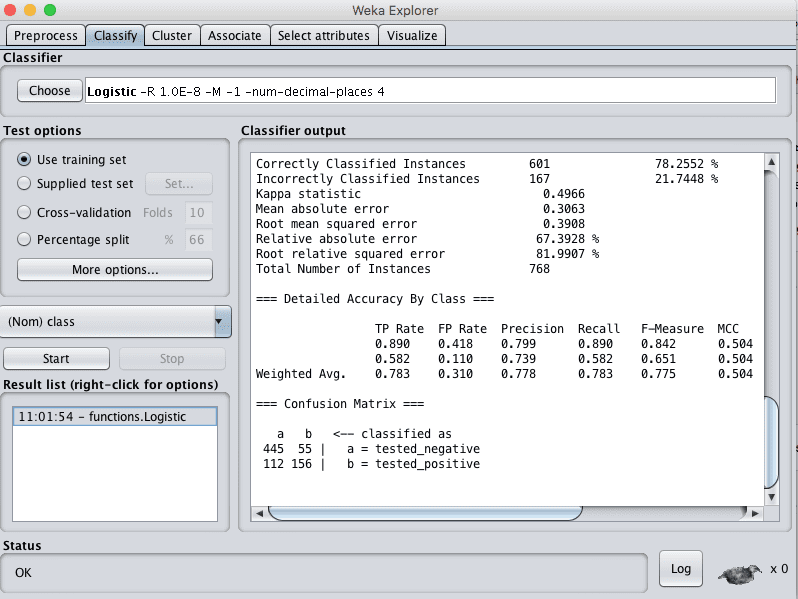
Weka Load Pima Indians Onset of Diabetes Dataset

4. Click the “Classify” tab to open up the classifiers.

5. Click the “Choose” button and choose “Logistic” under the “functions” group.

6. Select “Use training set” under “Test options”.

7. Click the “Start” button.



Weka Train Logistic Regression Model

This will train the chosen Logistic regression algorithm on the entire loaded dataset. It will also evaluate the model on the entire dataset, but we are not interested in this evaluation.

It is assumed that you have already estimated the performance of the model on unseen data using cross validation/percentage split as a part of selecting the algorithm you wish to finalise.

Now that we have finalised the model, we need to save it to file.

## 2. Save Finalized Model To File

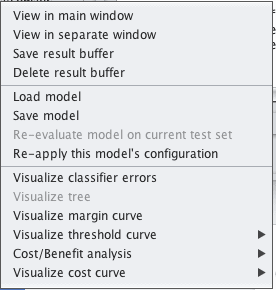
Continuing on from the previous section, we need to save the finalised model to a file on your disk.

This is so that we can load it up at a later time, or even on a different computer in the future and use it to make predictions. We will not need the training data in the future, just the model of that data.

You can easily save a trained model to file in the Weka Explorer interface.

1. Right click on the result item for your model in the “Result list” on the “Classify” tab.

2. Click “Save model” from the right click menu.



Weka Save Model to File

3. Select a location and enter a filename such as “logistic”, click the “Save button.

Your model is now saved to the file “logistic.model”.

It is in a binary format (not text) that can be read again by the Weka platform. As such, it is a good idea to note down the version of Weka you used to create the model file, just in case you need the same version of Weka in the future to load the model and make predictions. Generally, this will not be a problem, but it is a good safety precaution.

You can close the Weka Explorer now. The next step is to discover how to load up the saved model.

## 3. Load a Finalized Model

You can load saved Weka models from file.

The Weka Explorer interface makes this easy.

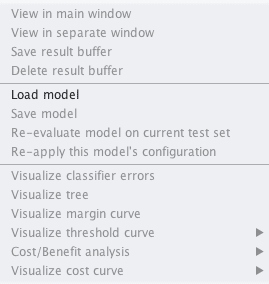
1. Open the Weka GUI Chooser.

2. Click the “Explorer” button to open the Weka Explorer interface.

3. Load any old dataset, it does not matter. We will not be using it, we just need to load a dataset to get access to the “Classify” tab. If you are unsure, load the data/diabetes.arff file again.

4. Click the “Classify” tab to open up the classifiers.

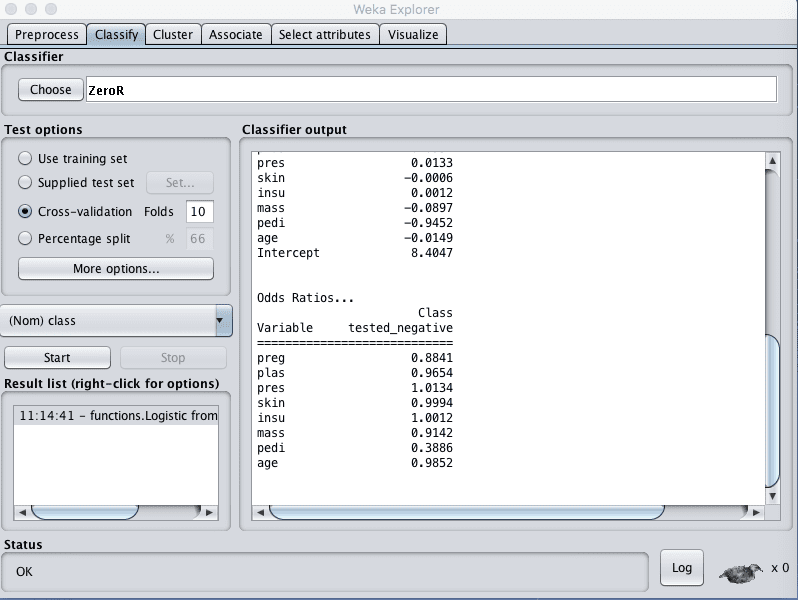
5. Right click on the “Result list” and click “Load model”, select the model saved in the previous section “logistic.model”.



Weka Load Model From File

The model will now be loaded into the explorer.

We can now use the loaded model to make predictions for new data.



Weka Model Loaded From File Ready For Use

## 4. Make Predictions on New Data

We can now make predictions on new data.

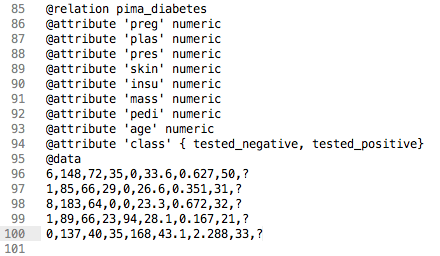
First, let’s create some pretend new data. Make a copy of the file “data/diabetes.arff” and save it as “data/diabetes-new-data.arff“.

Open the file in a text editor.

Find the start of the actual data in the file with the @data on line 95.

We only want to keep 5 records. Move down 5 lines, then delete all the remaining lines of the file.

The class value (output variable) that we want to predict is on the end of each line. Delete each of the 5 output variables and replace them with question mark symbols (?).

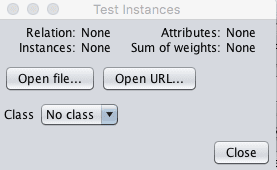


Weka Dataset For Making New Predictions

We now have “unseen” data with no known output for which we would like to make predictions.

Continue on from the previous part of the tutorial where we already have the model loaded.

1. On the “Classify” tab, select the “Supplied test set” option in the “Test options” pane.



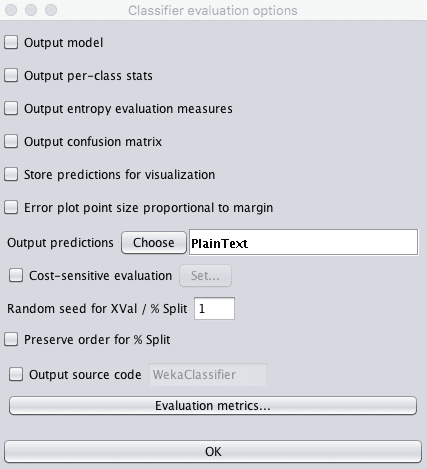
Weka Select New Dataset On Which To Make New Predictions

2. Click the “Set” button, click the “Open file” button on the options window and select the mock new dataset we just created with the name “diabetes-new-data.arff”. Click “Close” on the window.

3. Click the “More options…” button to bring up options for evaluating the classifier.

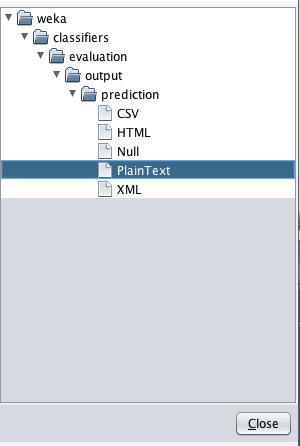
4. Uncheck the following information for now as we only want to see the predicted values, specifically:

* “Output model”
* “Output per-class stats”
* “Output confusion matrix”
* “Store predictions for visualization”



Weka Customized Test Options For Making Predictions

5. For the “Output predictions” option click the “Choose” button and select “PlainText”.

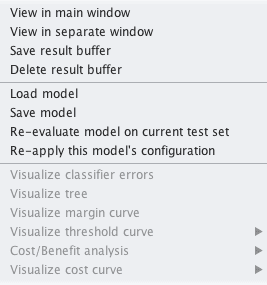


Weka Output Predictions in Plain Text Format

6. Click the “OK” button to confirm the Classifier evaluation options.

7. Right click on the list item for your loaded model in the “Results list” pane.

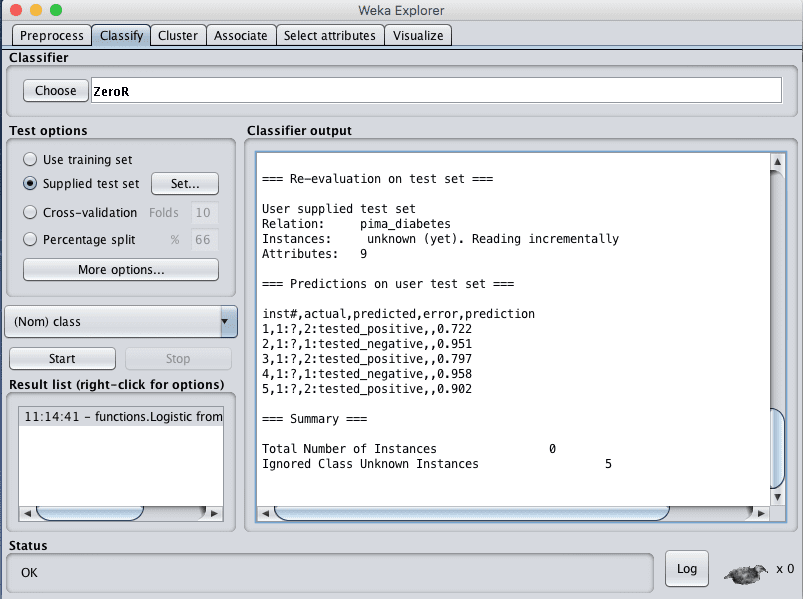
8. Select “Re-evaluate model on current test set”.



Weka Revaluate Loaded Model On Test Data And Make Predictions

The predictions for each test instance are then listed in the “Classifier Output” pane. Specifically, the middle column of the results with predictions like “tested\_positive” and “tested\_negative”.

You could choose another output format for the predictions, such as CSV, that you could later load into a spreadsheet like Excel. For example, below is an example of the same predictions in CSV format.



Weka Predictions Made on New Data By a Loaded Model

## Summary

In this post you discovered how to finalise your model and make predictions for new unseen data. You can see how you can use this process to make predictions on new data. Specifically, you have learned:

* How to train a final instance of your machine learning model.
* How to save a finalized model to file for later use.
* How to load a model from file and use it to make predictions on new data.

References

Jason Browlee, 2016, Machine Learning Mastery, <https://machinelearningmastery.com/>