Weka Data Analysis

Weka results output:
TP = true positives: number of examples predicted positive that are actually positive
FP = false positives: number of examples predicted positive that are actually negative
TN = true negatives: number of examples predicted negative that are actually negative
FN = false negatives: number of examples predicted negative that are actually positive

Weka Confusion Matrix:

	a	b	<-- classified as
actual a=0	TP	FN	
actual b=1	FN	TP	

Recall is the TP rate (also referred to as sensitivity)
what fraction of those that are actually positive were predicted positive? : TP / actual positives

Precision is TP / predicted Positive
what fraction of those predicted positive are actually positive?
precision is also referred to as Positive predictive value (PPV);
Other related measures used in classification include True Negative Rate and Accuracy:
True Negative Rate is also called Specificity. (TN / actual negatives)
1-specificity is x-axis of ROC curve: this is the same as the FP rate (FP / actual negatives)
what fraction of those that are actually negative were found to be positive?
(hopefully very low)

$$\text{Recall} = \frac{tp}{tp + fn}$$

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$$

F-measure

Main article: F1 score

A measure that combines precision and recall is the harmonic mean of precision and recall, the traditional F-measure or balanced F-score:

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

This is also known as the
F_1
measure, because recall and precision are evenly weighted.
Pasted from <http://en.wikipedia.org/wiki/Precision_and_recall>

Mean absolute error (MAE)
The MAE measures the average magnitude of the errors in a set of forecasts, without considering their direction. It measures *accuracy* for continuous variables. The equation is given in the library references. Expressed in words, the MAE is the average over the verification sample of the absolute values of the differences between forecast and the corresponding observation. The MAE is a linear score which means that all the individual differences are weighted equally in the average;

Root mean squared error (RMSE)
The RMSE is a quadratic scoring rule which measures the average magnitude of the error. The equation for the RMSE is given in both of the references. Expressing the formula in words, the difference between forecast and corresponding observed values are each squared and then averaged over the sample. Finally, the square root of the average is taken. **Since the errors are squared before they are averaged, the RMSE gives a relatively high weight to large errors. This means the RMSE is most useful when large errors are particularly undesirable.**

The MAE and the RMSE can be used together to diagnose the variation in the errors in a set of forecasts. The RMSE will always be larger or equal to the MAE; the greater difference between them, the greater the *variance* in the individual errors in the sample. If the RMSE=MAE, then all the errors are of the same magnitude

Pasted from
<http://www.eumetcal.org/resources/ukmeteocal/verification/www/english/msg/ver_cont_var/uos3/uos3_ko1.htm>

From the Weka Primer:

The kappa statistic measures the agreement of prediction with the true class -- 1.0 signifies complete agreement. **The following error values are not very meaningful for classification tasks, however for regression tasks e.g. the root of the mean squared error per example would be a reasonable criterion.** We will discuss the relation between confusion matrix and other measures in the text.

The confusion matrix is more commonly named *contingency table*. In our case we have two classes, and therefore a 2x2 confusion matrix, the matrix could be arbitrarily large. The number of correctly classified instances is the sum of diagonals in the matrix; all others are incorrectly classified (class "a" gets misclassified as "b" exactly twice, and class "b" gets misclassified as "a" three times).

The *True Positive (TP)* rate is the proportion of examples which were classified as class *x*, among all examples which truly have class *x*, i.e. how much part of the class was captured. It is equivalent to ***Recall*.** In the confusion matrix, this is the diagonal element divided by the sum over the relevant row, i.e. 7/(7+2)=0.778 for class *yes* and 2/(3+2)=0.4 for class *no* in our example.

The *False Positive (FP)* rate is the proportion of examples which were classified as class *x*, but belong to a different class, among all examples which are not of class *x*. In the matrix, this is the column sum of class *x* minus the diagonal element, divided by the rows sums of all other classes; i.e. 3/5=0.6 for class *yes* and 2/9=0.222 for class *no*.

The *Precision* is the proportion of the examples which truly have class *x* among all those which were classified as class *x*. In the matrix, this is the diagonal element divided by the sum over the relevant column, i.e. 7/(7+3)=0.7 for class *yes* and 2/(2+2)=0.5 for class *no*.

The *F-Measure* is simply 2*Precision*Recall/(Precision+Recall), a combined measure for precision and recall.

These measures are useful for comparing classifiers. However, if more detailed information about the classifier's predictions are necessary, *-p #* outputs just the predictions for each test instance, along with a range of one-based attribute ids (0 for none). Let's look at the following example. We shall assume soybean-train.arff and soybean-test.arff have been constructed via weka.filters.supervised.instance.StratifiedRemoveFolds as in a previous example.

Pasted from <http://weka.wikispaces.com/Primer>

Ex: Weka disease prediction results: Precision is the fraction of predicted diseased that are diseased, while recall is the fraction of diseased that are predicted diseased.

Kappa statistic	0.9403	<- agreement of prediction with true class
Mean absolute error	0.0309	<- not squared before averaging
Root mean squared error	0.1493	<- squared before averaging, so large errors have more influence
Relative absolute error	6.9047 %	<- Relative values are ratios, and have no units.

Absolute values have the same units as the quantities measured.
For example, 0.2314 grams, or plus or minus 0.02 mL.
Relative values are ratios, and have no units. The ratios are commonly expressed as fractions
(e.g. 0.562), as percent (fraction x 100, e.g. 56.2%),
as parts per thousand (fraction x 1000, e.g. 562 ppt), or as parts per million
(fraction x 106 , e.g. 562,000 ppm).

Scientific Reproducibility for Weka

Note: For reproducibility of experiments, copy the Scheme (the top line of the run information). This contains all the parameters and settings for the run:

Ex:
Scheme:weka.classifiers.functions.SMO -C 500.0 -L 0.0010 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.RBFKernel -C 250007 -G 10.0"

Right-click on this same line as displayed next to the “Choose” button to get information on each parameter for that learning method.