

How to install EBMC in WEKA?

May 5, 2014

Tutorial prepared by: Arturo Lopez Pineda

This tutorial covers the installation of the Efficient Bayesian Multivariate Classifier (EBMC) implemented as an external package for the Waikato Environment for Knowledge Analysis (WEKA).

For more information on these tools please refer to the original sources:

WEKA: <http://www.cs.waikato.ac.nz/ml/weka/>

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten. The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1, 2009

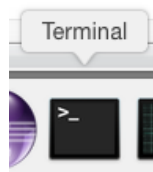
EBMC: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041321/>

Cooper, G. F., Hennings-Yeomans, P., Visweswaran, S., & Barmada, M. An efficient bayesian method for predicting clinical outcomes from genome-wide data. AMIA Annual Symposium Proceedings, 127–131, 2010

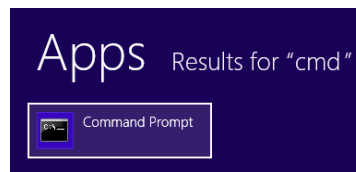
Step 1. Installing Java

Before installing Java, please verify which version your computer currently have installed. Enter `java -version` in the terminal or command line the following:

Mac OS X



Windows



The resulting answer to that command should be similar to:

```
java version "1.8.0_25"  
Java(TM) SE Runtime Environment (build 1.8.0_25-b17)  
Java HotSpot(TM) 64-Bit Server VM (build 25.25-b02, mixed mode)
```

In case you have previous versions you will have to install the latest. To install Java you will need the Java Runtime Environment (JRE) and the Java Development Kit (JDK). The JDK includes the JRE, so you do not have to download both. You can get JDK from:

JDK: <http://www.oracle.com/technetwork/java/javase/downloads/>

Please make sure to accept the license agreement, select your operating system and start the download:

Java SE Development Kit 8u25		
You must accept the Oracle Binary Code License Agreement for Java SE to download this software.		
<input type="radio"/> Accept License Agreement <input checked="" type="radio"/> Decline License Agreement		
Product / File Description	File Size	Download
Linux x86	135.24 MB	jdk-8u25-linux-i586.rpm
Linux x86	154.88 MB	jdk-8u25-linux-i586.tar.gz
Linux x64	135.6 MB	jdk-8u25-linux-x64.rpm
Linux x64	153.42 MB	jdk-8u25-linux-x64.tar.gz
Mac OS X x64	209.13 MB	jdk-8u25-macosx-x64.dmg
Solaris SPARC 64-bit (SVR4 package)	137.01 MB	jdk-8u25-solaris-sparcv9.tar.Z
Solaris SPARC 64-bit	97.14 MB	jdk-8u25-solaris-sparcv9.tar.gz
Solaris x64 (SVR4 package)	137.11 MB	jdk-8u25-solaris-x64.tar.Z
Solaris x64	94.24 MB	jdk-8u25-solaris-x64.tar.gz
Windows x86	157.26 MB	jdk-8u25-windows-i586.exe
Windows x64	169.62 MB	jdk-8u25-windows-x64.exe

After download double click the .dmg or .exe file and follow the wizard.

Mac OS X

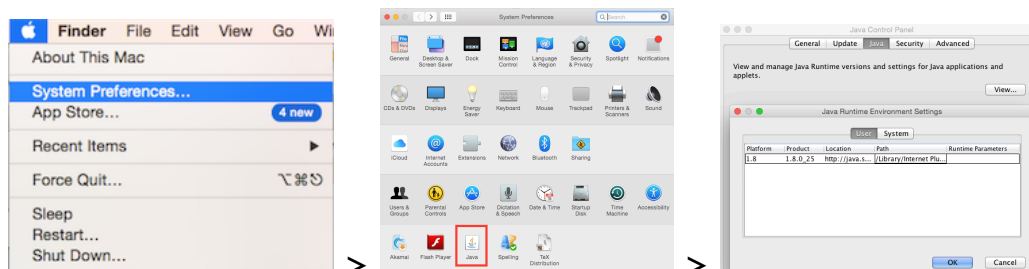


Windows




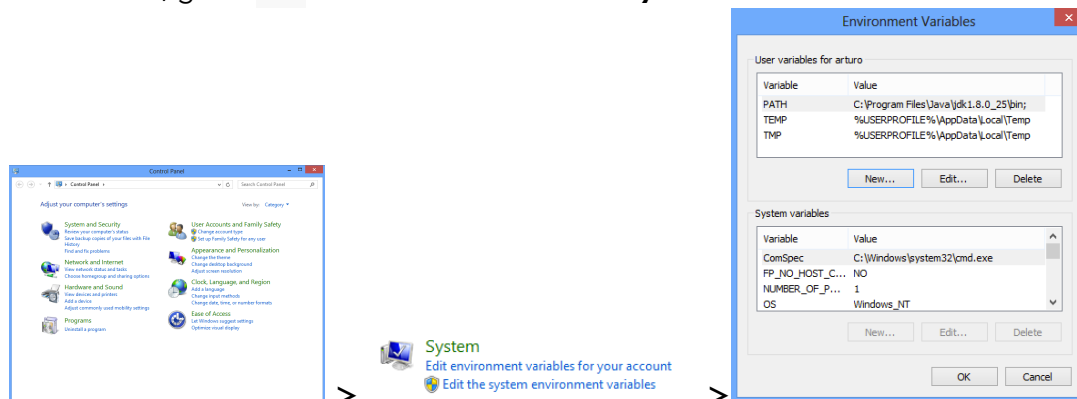
Now you have the latest version of Java installed. Let's make sure that your computer recognizes it.

In Mac OS X, go to **Apple > System Preferences > Java > Java (tab) > View**. You will be able to see the new version of JRE installed.



You can explore other tabs in the Java preferences for customization. After completion, go back to the Terminal and type `java -version` to make sure everything is working.

In Windows, go to  > **Control Panel** > **System** > **Environment Variables**



If there isn't a variable called `PATH`, create a new one. Otherwise, modify the existing one by adding to the end the path to the bin folder where Java JDK was installed. Remember to separate previous paths with a semicolon (;) mark. The most probable path is `C:\Program Files\Java\jdk1.8.0_25\bin`

Save the variable, click OK in the Environment Variables window, and reopen the Command Line to test your new path. Remember to use `java -version` to see whether the installation was successful.

Mac OS X

```
arturo ~  
Last login: Tue Oct 28 11:13:46 on ttys000  
-bash: dircolors: command not found  
[ arturo@tb-lx-wireless-pittnet-150-212-108-27 ~ ] java -version  
java version "1.8.0_25"  
Java(TM) SE Runtime Environment (build 1.8.0_25-b17)  
Java HotSpot(TM) 64-Bit Server VM (build 25.25-b02, mixed mode)  
[ arturo@tb-lx-wireless-pittnet-150-212-108-27 ~ ]
```

Windows

```
Microsoft Windows [Version 6.2.9200]  
(C) 2012 Microsoft Corporation. All rights reserved.  
C:\Users\arturo>java -version  
java version "1.8.0_25"  
Java(TM) SE Runtime Environment (build 1.8.0_25-b17)  
Java HotSpot(TM) 64-Bit Server VM (build 25.25-b02, mixed mode)  
C:\Users\arturo>
```

Step 2. Installing WEKA

You can download the latest version of WEKA from their website:

WEKA: <http://www.cs.waikato.ac.nz/ml/weka/downloading.html>

To be able to install external packages, please make sure to download version 3.7.10 or higher. From the developer version, look for the WEKA version appropriate for your computer. There is no need to install the version that contains Java. This was already covered in step 1 from this tutorial.

Be aware that WEKA stable version 3.6 does not allow the installation of external packages.

Windows:

- **Developer version**

This is the trunk of Weka and continues from the stable-3-6 code line. It receives both bug fixes and new features.

- **Windows x86**

Click [here](#) to download a self-extracting executable that includes Java VM 1.7 (weka-3-7-12jre.exe; 60.6 MB)

Click [here](#) to download a self-extracting executable without the Java VM (weka-3-7-12.exe; 33.2 MB)

After download double click the .exe file and follow the wizard.

Mac OS X:

- **Mac OS X**

Click [here](#) to download a disk image for OS X that contains a Mac application including Oracle's Java 1.7 JVM (weka-3-7-12-oracle-jvm.dmg; 92.9 MB)

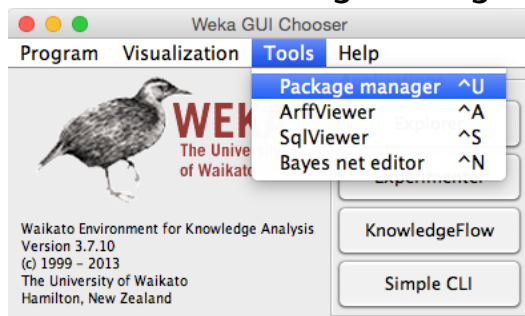
Click [here](#) to download a disk image for OS X that contains a Mac application compatible with Apple's Java 1.6 JVM (weka-3-7-12-apple-jvm.dmg; 41.3 MB)

After download double click the .dmg file, then drag and drop the WEKA program into the Applications folder.

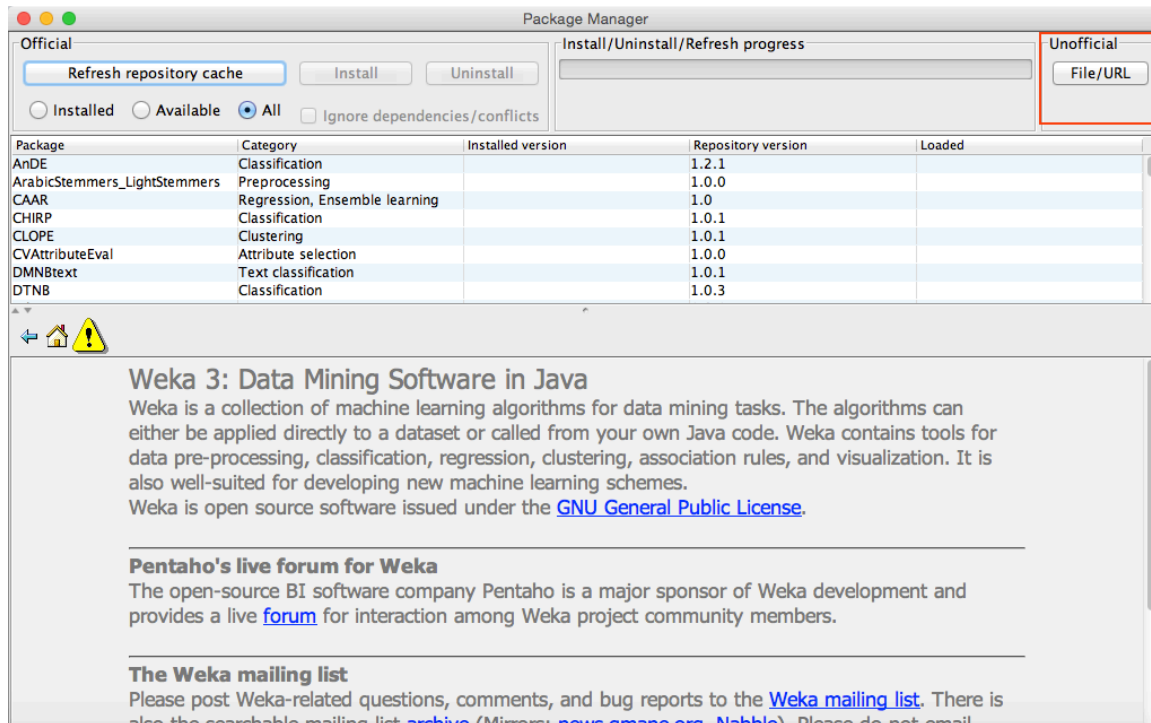
Step 3. Installing EBMC

Open WEKA by double clicking on the program icon (🍏). This will open a dialog window.

Go to **Tools > Package Manager**



And then click on **File/URL > Browse**



You will be prompted a new dialog window where you can search for the ebmc.zip file containing the EBMC package.

After installing you will have to close and reopen WEKA to be able to use EBMC.

Step 4. Using EBMC

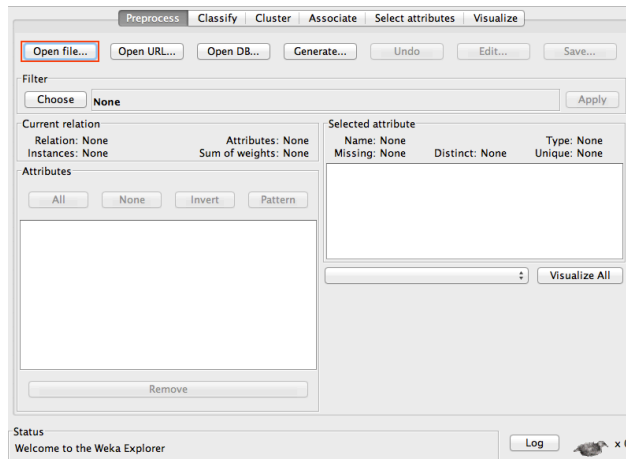
Double click on the **.arff** file that you want to classify, or open the WEKA explorer and select the .arff file.

Learn more about the ARFF file format in <http://www.cs.waikato.ac.nz/ml/weka/arff.html>

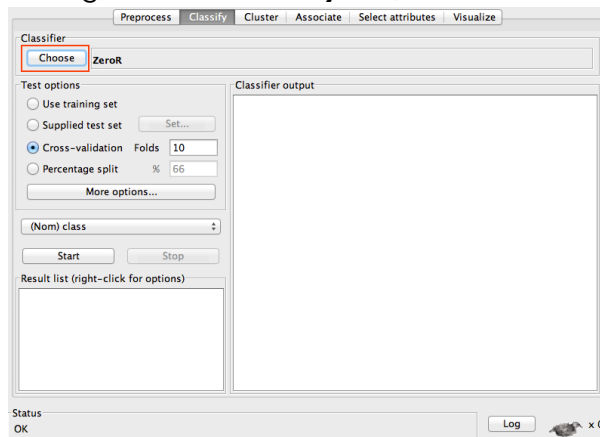
WEKA > Explorer



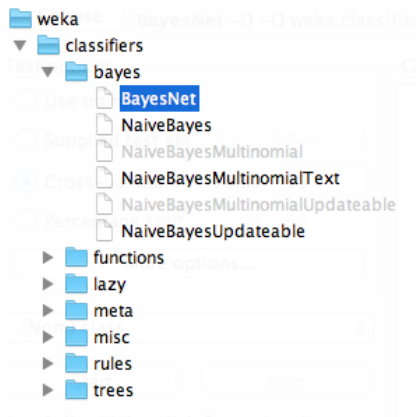
Browse the **.arff** file by clicking on **Open file...**



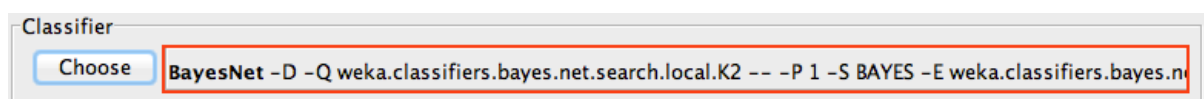
Then go to the **Classify** tab, and launch the **Choose** window:



From the list of classifiers select: **Bayes > BayesNet**

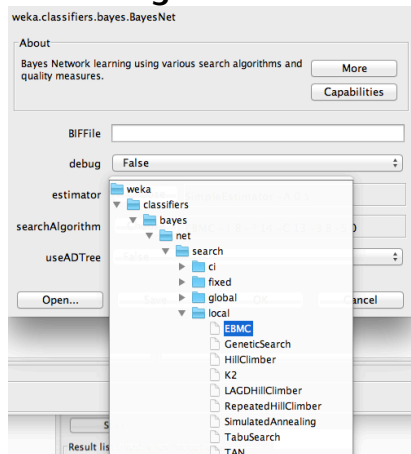


To set options in WEKA click on the gray box next to the Choose button of any element, for example:



Change the search algorithm options by selecting:

searchAlgorithm > classifiers > bayes > net > search > local > EBMC



Configure the EBMC options as needed. For more information on parameter selection refer to the original paper.

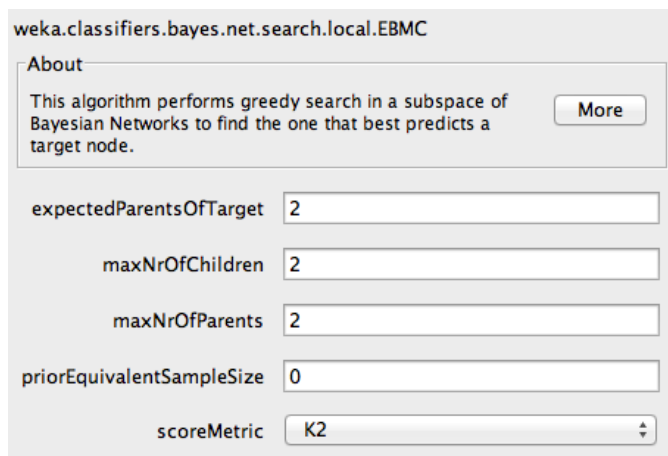
The available options are:

expectedParentsOfTarget -- Set the number of parents that the target node is expected to have.

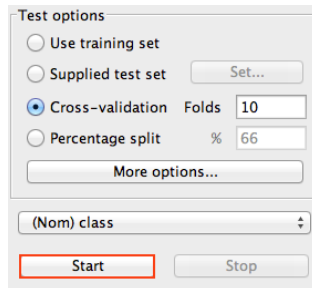
maxNrOfChildren -- Set the maximum number of children that each node can have.

maxNrOfParents -- Set the maximum number of parents that each node can have.

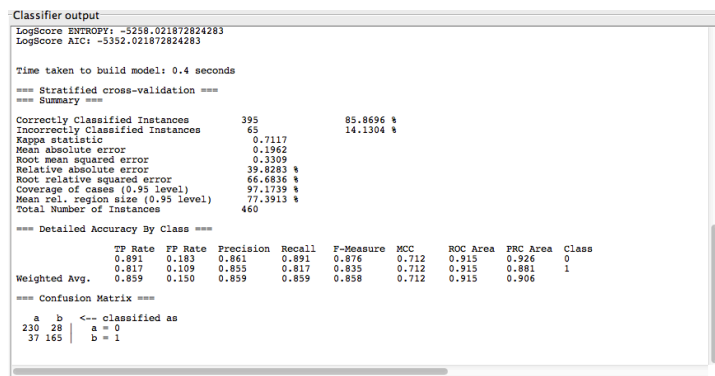
priorEquivalentSampleSize -- Set prior equivalent sample size used in the BDeu scoring measure



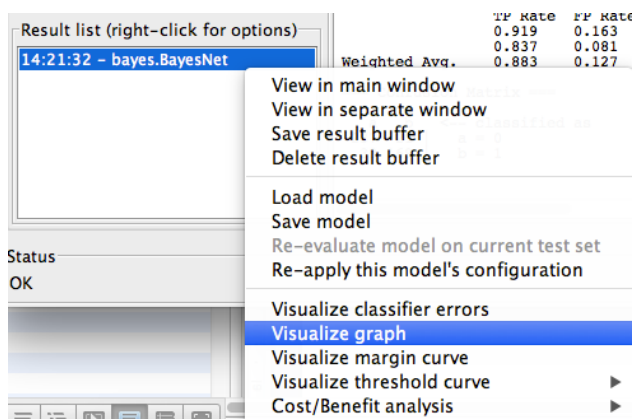
Finally, run the classification by clicking start depending on your experimental design (train/test, cross-validation, percentage split).



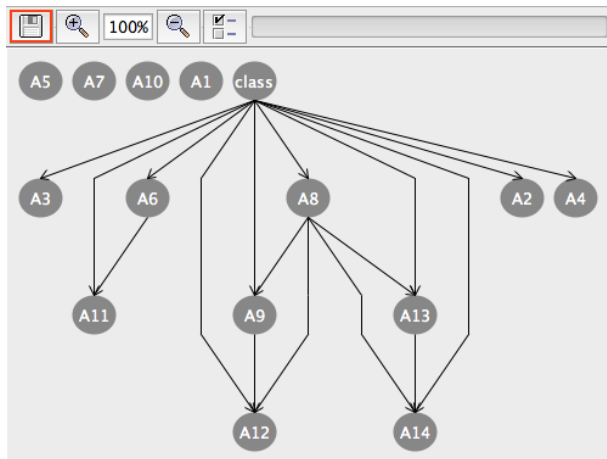
The results of the classification will be shown in the output panel, including precision, recall, AUC, among others.



To visualize the model right-click on the result list panel:
bayes.BayesNet > Visualize graph



The Bayesian network model can be seen in a new window. Double-clicking on any node shows the conditional probability table of that node. For the particular example shown below, the variables A5, A7, A10, and A1 were discarded by the EBMC search, and therefore they do not form part of the network.



The model can be saved to the chosen destination by clicking on the save button and selecting **XML BIF files** format.

This visualization option is only available under the BayesNet classifiers; therefore, it will not show in any of the Naïve Bayes classifiers.

Step 5. Tracking the EBMC search strategy

WEKA's GUI does not show the EBMC search strategy explicitly. However, using the command line there is a way to visualize the scoring that all models obtained and how EBMC added or deleted nodes from the network.

First, find the **weka.jar** file in your system directory. Most probably is in:

Windows: c:\Program Files (x86)\Weka-3-7\

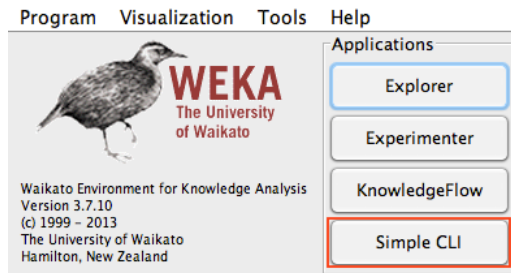
Mac OS X: /Applications/weka-3-7-12-apple-jvm.app/Contents/Resources/Java

Open the terminal or command line. If you have forgotten how to open it, check the first part of STEP 1 in this tutorial.

Use the command `cd` (change directory) to change the command/terminal prompt to that specific directory and then type:

```
java -jar weka.jar
```

This will open the WEKA GUI that will allow you to run the same thing as in STEP 4. As an alternative you can also click on the Simple CLI to launch an additional window that will show the same as the terminal/command line.



This time, when clicking on the run button the command line will also show the underlying search that EBMC is performing in each step. For the dataset `australia.arff` the output for the configuration (`expectedTargets=8`, `maxParents=14`, `maxChildren=13`) is:

```
score: -330.14161215743746
starting search for an additional predictor
  current new predictors: A8      score:-207.0820611333828
  current new predictors: A8 A9   score:-191.2097497585141
  current new predictors: A8 A9 A12 score:-186.08926296992448
  current new predictors: A8 A9 A12 A13 score:-182.2401175928469
  current new predictors: A8 A9 A12 A13 A14 score:-180.10676105195324
score: -180.10676105195324
starting search for an additional predictor
  current new predictors: A7      score:-175.6581719215472
  current new predictors: A7 A4   score:-173.30315767807923
score: -173.3031576780793
starting search for an additional predictor
  current new predictors: A6      score:-171.8521413329735
  current new predictors: A6 A11  score:-170.09089190482163
score: -170.09089190482163
starting search for an additional predictor
  current new predictors: A2      score:-168.098025638984
score: -168.098025638984
starting search for an additional predictor
  current new predictors: A3      score:-168.09204778061275
score: -168.09204778061275
starting search for an additional predictor
Pruning node A9 away from node A13
Pruning node A12 away from node A13
Pruning node A9 away from node A14
Pruning node A12 away from node A14
Pruning node class away from node A7
Pruning node A7 away from node A4

=====
MODEL:
A1 (2)
A2 (2) --> class
A3 (2) --> class
A4 (3) --> class
A5 (14)
A6 (9) --> class
A7 (2)
A8 (2) --> class
A9 (2) --> class, A8
A10 (3)
A11 (2) --> class, A6
A12 (3) --> class, A8, A9
A13 (2) --> class, A8
A14 (2) --> class, A8, A13
class (2)
```