MixDB manual:

1) Generate decoy database:

Run the script DecoySeqGenerate to append a decoy database to your fasta file as follow:

usage: java -Xmx1000M -jar DecoySeqGenerate.jar <fasta file> <outfile>

2) To run MixDB:

Run MixDB as follow:

usage: java -Xmx2000M -jar MixDB.jar <database> <spectraFile> <parentmass tolerance> <fragment mass tolerance> <outfile> This will search the sequence database and find the best *pair* of peptide that best matched to the query spectrum Precursor mass tolerance is in unit Da. Usually one should use a relative large tolerance like 3Da to allow for the identification of mixture spectra even if query is high accuracy MS data MixDB accept input spectrum file in mzXML format, if you have spectrum file in other format you can try convert your spectrum file to mzXML format using proteowizard: <http://proteowizard.sourceforge.net/>

3) Running SVM classification

After the search, MixDB use a SVM to determine whether a match is significant,

SVM classification is done using the svm-light package, please go download the binary here: <http://svmlight.joachims.org/>

And put the appropriate binary: svm\_learn and svm\_classify in the svm\_light\_linux or svm\_light\_windows folder depending on your system

Use the MixDBFilter script to filter the search results as follow:

usage: java -Xmx1000M -jar MixDBFilter.jar <mixdb results> <output> <fdr>

note, please use MixDBFilterWin.jar if you are under windows system to call the appropriate svm binary

Output:

Outputs are in tab-delimited format, each column has the following meanings:

We denote M as the query spectrum and A and B as the pair of peptide best matched to M. In case of

|  |  |
| --- | --- |
| Column | Content |
| 1 | Spectrum file name |
| 2 | Spectrum scan number |
| 3 | Peptide |
| 4 | Protein |
| 5 | Raw combine score |
| 6 | Raw score (one peptide) |
| 7 | Explained Intensity |
| 8 | % b |
| 9 | %y |
| 10 | B ions series |
| 11 | Y ions series |
| 12 | svm-score for matches, high score means at least peptide A significant match to M |
| 13 | svm-score for mixture matches, higher score mean both A and B are significant matches to M |
| Additional notes: | When a matches is classified as a mixture matches where two peptides are matched to a query spectrum, column 3-11 will contain two entries per column separated by a symbol '!' with each entry correspond to each peptide match |