# Random Bits Forest

Author: Yi Wang
9/Oct/2015

**Hardware Requirement:**

The software requires modern CPUs with **SSE** instructions.

**Usage:**

**rbf** *[options]* ***trainX trainY testX testYhat***

**Format:**

trainX and testX should be comma delimited or colon delimited or tab delimited text files. The data is a dense matrix without header. Each line represents a sample and each column represents a variable. Missing value is define as |X|>=1e15. Missing value will be replaced with mean value automatically.

trainY should contain one column. Each line is one response/target value for each sample. The output file testYhat is similar to trainY. Each line is the predicted value for each test sample.

**Options:**

**-b**

Number of random bits used. The default setting (8192 bits) balances well on various datasets and the parameter is usually not tuned.

**-i**

Number of independent boosting chain. Larger value is equivalent to larger regularization. For small sample size dataset, you can set a big value (but should be less than –b/32).

**-c**

Number of candidates for each random bit. We generate -c random neural network candidates, and pick the best one as one random bit. Larger value take more time but produces better result.

**-1**

The number of features involved in first layer neural network. This parameter should be tuned with small integers (2-9). Usually 2 or 3 are choice.

**-2**

The number of node in second layer neural network. This parameter is usually 2.

**-n**

Number of trees in the forest. Larger value is slightly better but take more time.

**-s**

Sample bootstrap fold. Smaller value (<1) is corresponding to more regularization. Larger value takes more time.

**-f**

Feature bootstrap fold. Smaller value (<1) is corresponding to more regularization. Larger value takes more time.

**Enjoy!**