

Description:

GetAUC.pl obtains the AUC, ROC curve, precision at several recall values for various GO categories, and the median, average and standard deviation for all the measures across the GO categories evaluated.

Usage:

```
getAUC.pl goMatrix_file goCategories_file genes_file [verbose-level]
```

Arguments:

- goMatrix_file .- a tab-delimited text file with gene IDs as rows and GO categories as columns indicating whether a gene is annotated in a given GO category (0 – no annotated, 1 – annotated).
- goCategories_file .- a text file listing (one per line) the GO categories in the goMatrix_file to be evaluated. Each line should at least contain the GO ID (e.g., GO:0004567)
- genes_file .- a text file listing (one per line) the gene IDs of interest. Each line should contain the gene ID at the beginning of the line.
- verbose-level .- an integer between 0 and 2 indicating the amount of printing the program does (default 0).

Additional Input:

getAUC looks for a file per GO category listed in the goCategories_file with the scores (ranging from 0 to 1) per gene (one gene per line); i.e., geneID score. The score file can be delimited by any space character or comma. The score file name must be “GOID”_pred.txt (e.g., GO:0004567_pred.txt).

Output:

- getAUC outputs to STDIN a line per GO category containing the GO ID, number of positive instances, number of negative instances, precision at various recalls, AUC, and AUC_50. The precision at each specific recall is the precision obtained when reaching that recall (i.e., highest precision associated with each recall value). The last lines printed indicate the median, average and standard deviation of all the measures across the GO categories evaluated.
- A tab-delimited text file per GO category named ROC.GOID (e.g., ROC. GO:0004567) containing the points of the ROC curve. Each line contains FPR, TPR and score.

References:

- The algorithm used to calculate the AUC and the ROC curve is the one described in: Tom Fawcett (2003). "ROC Graphs: Notes and Practical Considerations for Researchers", HP Labs Tech Report HPL-2003-4. Available from: <http://www.purl.org/net/tfawcett/papers/ROC101.pdf>
- Interpolation between two points in Precision-Recall space is done as described in: Jesse Davis & Mark Goadrich, “The Relationship Between Precision-Recall and ROC Curves”, ICML 2006. Available from: <http://www.cs.wisc.edu/~richm/articles/davisgoadrichcamera2.pdf>

Contact:

If you find a bug or have a question/comment, please send an email to Lourdes Pena Castillo (lourdes (dot) pena (at) gmail (dot) com)