

FaST-LMM.Py v2.02 (Atmosphere Images Tutorial)

Introduction

This tutorial is intended to introduce new users to the FaST-LMM software for GWAS analysis. This Atmosphere image is publicly available under the name *FaST-LMM.Py v2.02*.

All of the necessary Python modules are already installed on this instance, so you can get started analyzing right away!

 This is a tutorial for FaST-LMM as a distinct Atmosphere image. For FaST-LMM as a step in the Validate Workflow see [here](#).

 **Learn about allocations**
Learn about CyVerse's allocation policies [here](#).

- [Introduction](#)
 - [About the software](#)
- [Accessing FaST-LMM](#)
 - [Testing FaST-LMM](#)
 - [Trying out your data](#)
 - [Running example data](#)
 - [Explanation of the code line](#)
- [Additional information](#)

About the software

FaST-LMM (*Factored Spectrally Transformed Linear Mixed Models*) is a GWAS analysis tool designed for large data sets. Normally, running a linear mixed model on a dataset is thorough, but very computationally demanding and may not even work on especially big data sets. FaST-LMM, however, changes things by reducing the runtime needed to produce such a model. Normally, when dealing with SNPs, a genetic similarity matrix is formed. FaST-LMM works by obtaining the spectral decomposition of this similarity matrix without actually computing the matrix itself. This decomposition is then used to test all SNPs in the data set for statistical significance. Such a method allows for proportionally smaller computation time in contrast to other programs. For a more in-depth explanation, see the paper from Microsoft Research [here](#).

Accessing FaST-LMM

To use FaST-LMM via VNC Viewer, follow these simple steps:

1. Launch a new instance of FaST-LMM.Py v2.02 from Atmosphere, and access it using VNC.
2. Once you have access to the instance, open up the terminal by either clicking the black icon at the bottom of the screen, or by going to Application > Accessories > Terminal.
3. Begin coding!

Testing FaST-LMM

The first thing to do is test that the python aspects of the image are working correctly. Where you see "sudo" written into the commands below is where the commands are being performed as the root user.

1. First, open up your terminal.
2. Change your directory to the feature_selection folder:

```
cd /usr/FaST-LMM-master/fastlmm/feature_selection
```

3. Run the test.py file using the following code:

```
sudo python test.py
```

```
Terminal - gpcarpen@vm64-134: /usr/FaST-LMM-master/fastlmm/feature_selection
File Edit View Terminal Go Help
gpcarpen@vm64-134:~$ cd /usr/FaST-LMM-master/fastlmm/feature_selection
gpcarpen@vm64-134: /usr/FaST-LMM-master/fastlmm/feature_selection$
test
```

4. You should see a large amount of code flash on the screen. This is normal. The whole testing process should take 7-8 minutes, and once it is finished you will see OK at the bottom of the screen.

```
Terminal - gpcarpen@vm64-134: /usr/FaST-LMM-master/fastlmm/feature_selection
File Edit View Terminal Go Help
mse: 0.526583
k: 16
time taken for k=16: 0.558046102524
mse: 0.526502
k: 32
time taken for k=32: 0.94878411293
mse: 0.485716
k: 64
time taken for k=64: 2.13907313347
mse: 0.490465
time taken for fold: 4.70406007767
using MSE to select best_k
best_k: 32, best_mix: 0.678657, best_delta: 0.701484
final scan using LR
number of snps selected: 32
best_k: 32
best_mix: 0.678656620586
best_delta: 0.701484427044
.
-----
Ran 18 tests in 394.584s
OK
gpcarpen@vm64-134: /usr/FaST-LMM-master/fastlmm/feature_selection$
```

5. You are now able to test your data!

Trying out your data

 Depending on the size of your dataset, you may wish to use the Stampede system for your data analysis. A FaST-LMM application on Stampede is forthcoming, or you may upload the software manually and use it that way.

One thing to keep in mind is that the actual FaST-LMM program is a C-based file. The Python-compatible functions from FaST-LMM, while important, are technically extensions of this main program. The most important Python-compatible functions are as follows:

- SNP Selection (FaST-LMM-Select)
- SNP-set testing (as opposed to single SNP testing used by the main program; FaST-LMM-Set)
- Tests for Epistasis

If you wish to try out any of these Python functions, please consult the documentation located in the usr directory under FaST-LMM-Docs. This document can be accessed in the image by minimizing the terminal and accessing through the file path.. File manager > usr > Fast-Lmm-Docs.

The remainder of this tutorial will be focused strictly on the main FaST-LMM program and its output(s).

The main program may be accessed simply by typing in **fastlmmc** in the terminal as it is saved into the `usr/bin/` folder.

FaST-LMM uses four primary input files:

1. SNP files to be tested
2. SNP data used to determine genetic similarities between individuals (this file can be different from the first)
3. A file containing phenotype data.
NOTE: This file should be in PLINK phenotype format with at least three columns for **familyID**, **individualID**, and **phenotypeValue** delimited by tab or whitespace
4. A covariate file (optional).
NOTE: This file should have at least three tab-delimited columns representing **familyID**, **individualID**, and **covariateValue**

The first two SNP files must be in PLINK format? (PED/MAP, BED/BIM/FAM, TPED/TFAM). The input flags you can use are as follows:

- **-file** : Denotes the file name for the PLINK .ped/.map files*
- **-bfile** : Denotes the name for PLINK .bed/.bim/.fam files*
- **-tfile** : Denotes the name for PLINK .tped/.tfam files*
- **-pheno** : Denotes the name of the phenotype file (including extension)
- **-covar** : Denotes the name of the covariate file (including file extension)
- **-out** : The name of your final output file, which is output into the same directory as your program and data unless otherwise specified

These are the bare minimum options needed to run FaST-LMM; however, some other options for considerations are...

- **-verboseOutput** : use this flag to show more complex and detailed output; does not require a file to be named
- **-extract** : A SNP filtering option used in conjunction with FaST-LMM-Select. FaST-LMM will only use SNPs listed in the input file for analysis
- **-pValuePrintThreshold** : Restricts the output file to only include SNPs with a p-value less than or equal to the specified threshold

Running example data

1. To run example data, first change the working directory:

```
cd /usr/FaST-LMM-master/fastlmm/feature_selection/examples
```

2. Then you can run the command line options for FaST-LMM:

```
fastlmmc -verboseOutput -bfile toydata -fileSim toydata -pheno toydata.phe -covar  
toydata.cov -out ~/Desktop/MyResults.csv -pValuePrintThreshold 0.05
```

Explanation of the code line

1. `fastlmmc` – This flag is the main executable
2. `-verboseOutput` – This flag triggers verbose mode; gives more detail for output
3. `-bfile` – This flag indicates the binary PLINK file set to use in the analysis (BED/BIM/FAM set). DOES NOT INCLUDE FILE EXTENSION.
4. `-fileSim` (can also be `bfileSim` or `tfileSim`) – This flag indicates the PLINK file set to use for computing the genetic similarity matrix. Can be the same file set as the previous command. DOES NOT INCLUDE FILE EXTENSION.

`fileSim` – means using a PED/MAP set

`bfileSim` – means using a BED/BIM/FAM set

`tfileSim` – means using a TPED/TFAM set

5. `-pheno` – This flag indicates the phenotype file for the file set to the analysis. THIS DOES INCLUDE FILE EXTENSION.
6. `-covar` – This flag indicates the covariate file and is optional. THIS DOES INCLUDE FILE EXTENSION.
7. `-out` – This flag indicates the name and location of the output files. The file extension will automatically output to `txt` unless otherwise specified, `.CSV` is more efficient for later certain data analysis though.
8. `-pValuePrintThreshold` – This flag tells the program to print only p values < 0.05 and is optional.

This will output to the desktop on your atmosphere image and the output will look like [this](#).



NOTICE

Make sure that you either have all of your data you want to analyze in the same folder as fastlmmc or have specified the correct path to your files!

Additional information

If you want more information on how to run FaST-LMM, documentation can be found here:

```
/usr/bin/FastLmm_Docs
```

and demo data can be found here:

```
/usr/FaST-LMM-master/fastlmm/feature_selection/examples
```