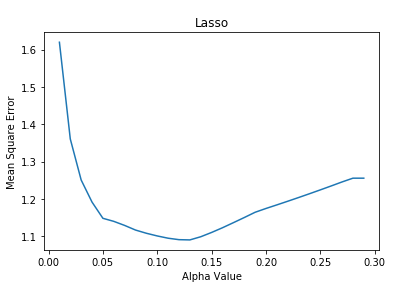
CSE 527 HW 1

*Chenchao Xu*

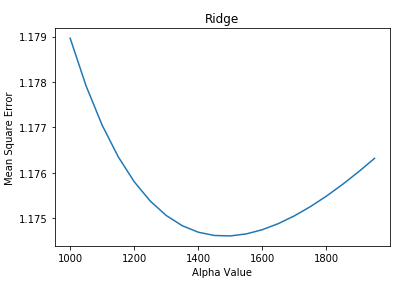
**2. Multi-marker model for QTL analysis (code\_q2.py)**

**a.**



min(lasso\_mse) = 1.090

**b.**



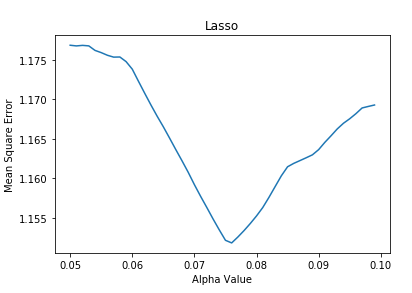
min(ridge\_mse) = 1.175

MSE with L2 is much higher that it with L1.

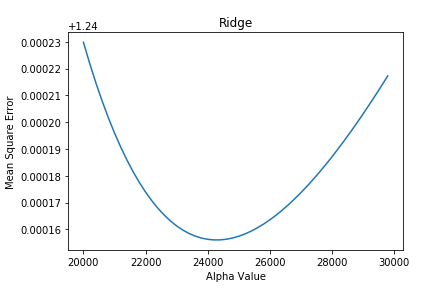
**c.**

train set : validation set = 2:1; randomly splited

And then use both training set and validation set to train the model for test set.



alpha\_lasso = 0.076



alpha\_ridge = 24200

On test set:

lasso\_mse\_test = 1.120; ridge\_mse\_test = 1.234; Lasso does better

parameters:

lasso: 0.076, not so far; ridge: 24200, much bigger than that in b

**3. Implementation of the EM-based haplotype reconstruction algorithm (code\_q3.py)**

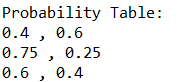
**a.**

hidden variables: haplotype states of individuals

parameters: haplotype frequencies

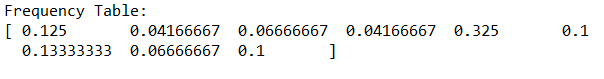
**b.**

Use current frequency estimates to replace ambiguous genotypes with fractional counts of phased genotypes

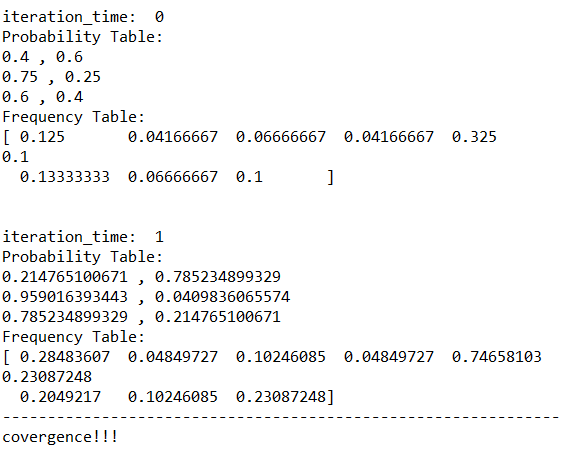


**c.**

Estimate frequency of each haplotype by counting



**d.**



The last iteration\_time is 1.