**Approach 1**

Prior vector on the gene and GO clusters in GAIL.

Compromise between hand curated prior and GAIL cluster.

**Approach 2**

Basically a regression.

**Y = X**

Estimate the probability that each gene belongs to each pathway.

Use classification, which is a supervised learning. We will do “semi-supervised” approach, where we guide some elements but not others.

Use some kind of variable selection to see which go twrms are important for extending pathway information.

High dimensional problem.

KEGG genes are the hand curated ones.

Ridge: shrinkage, no selection

Lasso: shrinkage and selection. Problem with collinearity

Elastic Net: compromise between LASSO and ridge

**Approach 3**

From KEGG we know the membership.

For each unannotated gene (one by one) check similarity to KEGG annotation genes.

**Approach 4**