

Readme for the Programs/Data used for the Sparse SVD paper

1. Main function (R/Matlab): ssvd.R (ssvd.m), thresh.R (thresh.m)
2. Input variables:
 - (a) X= argument (n x d matrix)
 - (b) threu = type of penalty (thresholding rule) for the left singular vector
 - 1 = (Adaptive) LASSO (default)
 - 2 = hard thresholding
 - 3 = SCAD
 - (c) threv = type of penalty (thresholding rule) for the right singular vector
 - 1 = (Adaptive) LASSO (default)
 - 2 = hard thresholding
 - 3 = SCAD
 - (d) gamu = weight parameter in Adaptive LASSO for the left singular vector
nonnegative constant (default = 0, LASSO)
 - (e) gamv = weight parameter in Adaptive LASSO for the right singular vector
nonnegative constant (default = 0, LASSO)
 - (f) u0, v0 = initial values of left/right singular vectors (default = the standard SVs)
 - (g) merr = threshold to decide convergence (default = 10^{-4})
 - (h) niter = maximum number of iterations (default = 100)
3. Output:
 - (a) u = left sparse singular vector
 - (b) v = right sparse singular vector
 - (c) iter = number of iterations to achieve the convergence

4. Lung cancer data

The data.txt file is a 12625 x 56 matrix; each column (row) of X corresponds to the individual case (gene).

Among the 56 cases,

Columns 1~20: pulmonary carcinoid samples (Carcinoid);

Columns 21~33: colon cancer metastasis samples (Colon);

Columns 34~50: normal lung samples (Normal);

Columns 51~56: small cell carcinoma samples (SmallCell).