

To set the parameters of *Hypergene* for various classification tasks, we do a k-fold cross validation on training set. The result for Rosetta dataset with 213 important genes is

Cross validation on Rosetta dataset with 231 important genes

ROC	$\rho = 0.0001$	$\rho = 0.001$	$\rho = 0.01$	$\rho = 0.1$	$\rho = 1$
$\alpha = 0.3$	0.79116	0.81964	0.81806	0.8217	0.8217
$\alpha = 0.5$	0.80554	0.8126	0.82004	0.82054	0.82054
$\alpha = 0.7$	0.81192	0.81578	0.82012	0.81958	0.81958

The parameters which has the best overall performance on training set is $\alpha = 0.3$ and $\rho = 0.1/1$. Actually, the two sets of parameters also have the same performance on test set.

We do the cross validation again with 500 top genes ranked by correlation coefficient. The result is

Cross validation on Rosetta dataset with 500 top ranked genes

ROC	$\rho = 0.0001$	$\rho = 0.001$	$\rho = 0.01$	$\rho = 0.1$	$\rho = 1$
$\alpha = 0.3$	0.84616	0.87824	0.88062	0.88066	0.88066
$\alpha = 0.5$	0.85406	0.87816	0.87824	0.87822	0.87822
$\alpha = 0.7$	0.86374	0.87536	0.87534	0.87574	0.87574

The best parameters are the same and they also have the same performance on test set.

We also do cross validation on Wang's dataset with 700 top genes ranked by correlation coefficient. The result is

Cross validation on Wang dataset with 700 top ranked genes

ROC	$\rho = 0.0001$	$\rho = 0.001$	$\rho = 0.01$	$\rho = 0.1$	$\rho = 1$
$\alpha = 0.5$	0.5645	0.5551	0.5542	0.5546	0.5542

(To save time, we just test the case $\alpha = 0.5$). The best parameter is $\alpha = 0.5$ and $\rho = 0.0001$. We use this set of parameter on test set.