Initial Results of Probit Regression on Flu Data

Minhua Chen

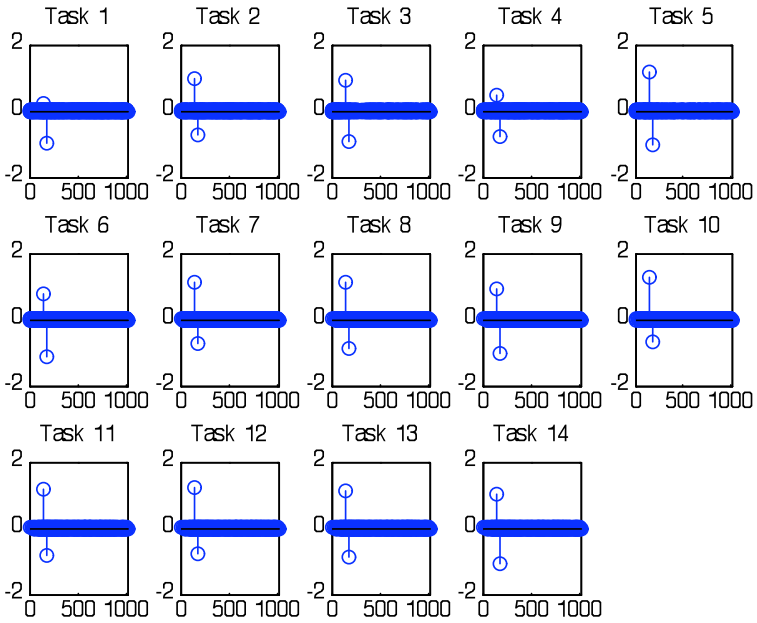
11.30.2008

Method

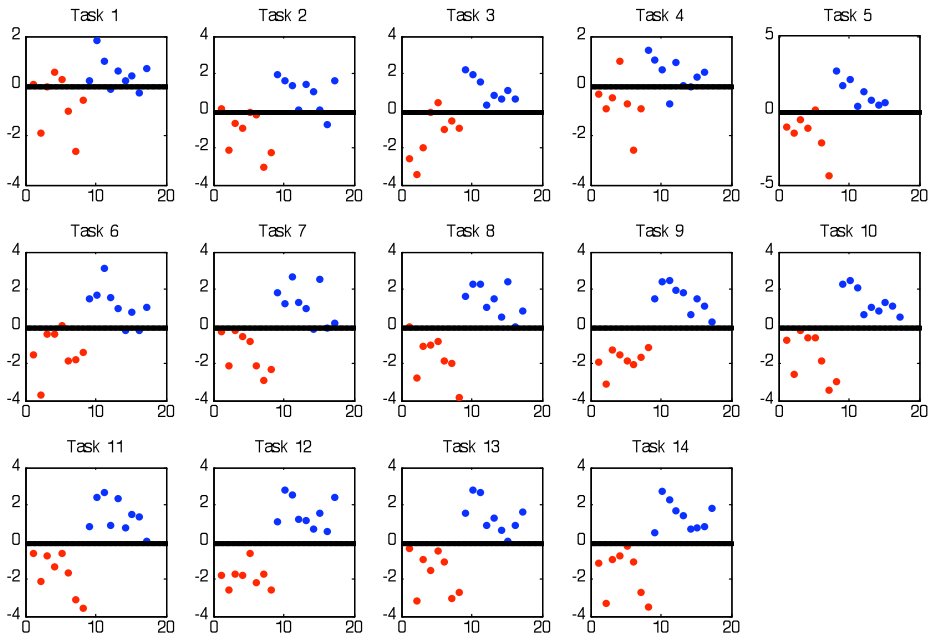
- Use data from 5 hrs to 108 hrs as training data. Each time point is a task, with 14 tasks in all.
- Two label schemes are tried: modified label (where subject 13 and 15 are treated as sick people), and the original label.
- 0.1T is around 5 hrs to 9.8 hrs, 0.8T is around 40 hrs to 78.4 hrs.
- The original 12023 genes are reduced to 1000 genes in advance, using F-score on late time data.
- Bayesian LASSO and Bayesian Elastic Net model are tried. Bayesian LASSO result is similar as the student-t version. Bayesian Elastic Net can select more genes then Bayesian LASSO.
- Genes selected using the two label schemes are totally different.

Time	baseline	0 hrs	5 hrs (Task1)	12 hrs	21.5 hrs	29 hrs	36 hrs	45.5 hrs	53 hrs
Samples	9 + 7	10 + 7	10 + 7	10 + 7	10 + 6	9 + 7	8 + 7	10 + 7	10 + 7
Time	60 hrs	69.5 hrs	77 hrs	84 hrs	93.5 hrs	101 hrs	108 hrs (Task14)		
Samples	10 + 7	10 + 7	10 + 7	10 + 7	10 + 7	10 + 7	10 + 7		

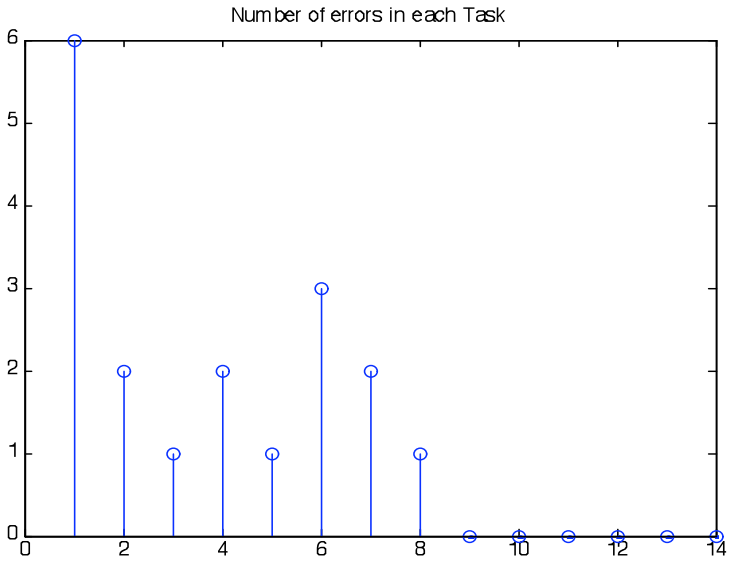
9 + 7 means 9 healthy samples and 7 influenza infected.



Regression coefficients for each task



Regression response for each task

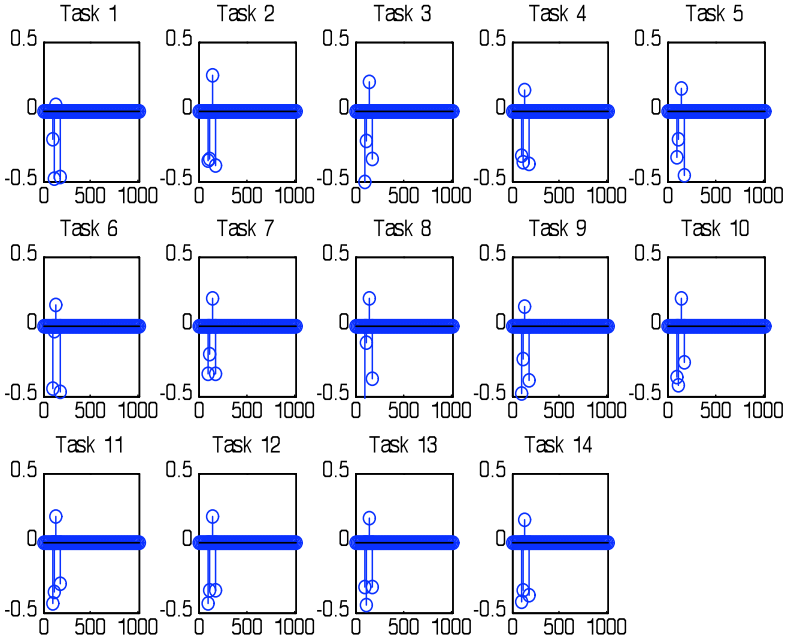


Number of errors in each task

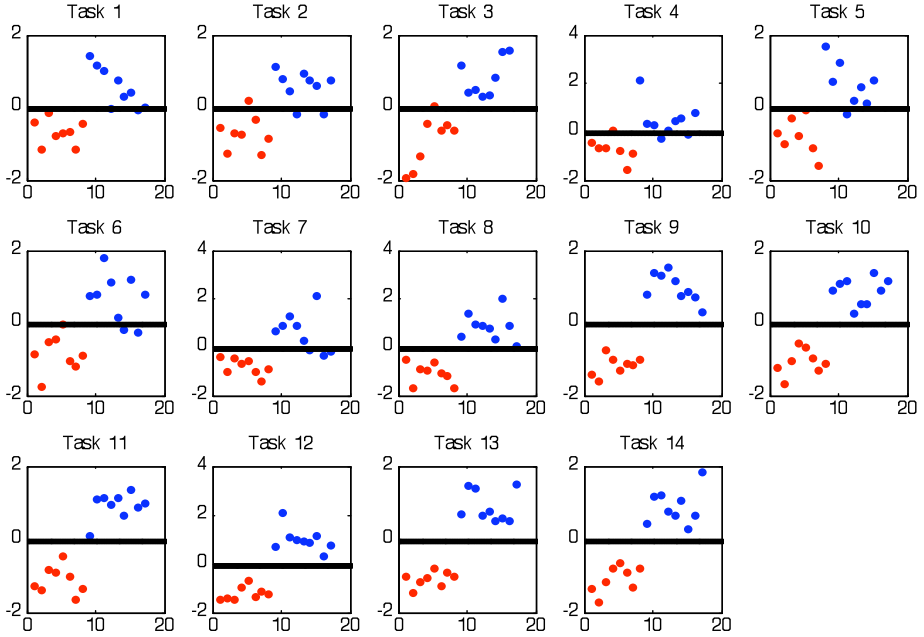
Bayesian LASSO (Modified label)

Gene identified (2):

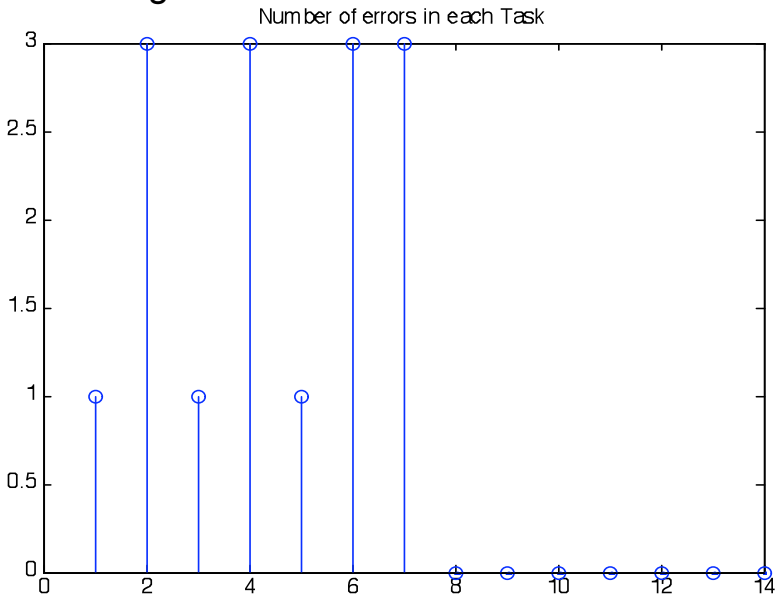
'LILRB2' '10288_at'
'HLA-DOB' '3112_at'



Regression coefficients for each task



Regression response for each task

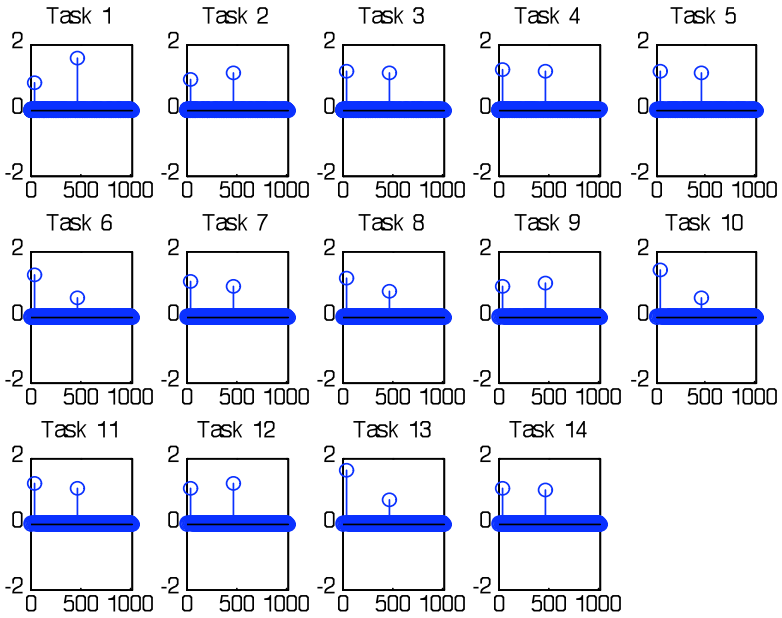


Number of errors in each task

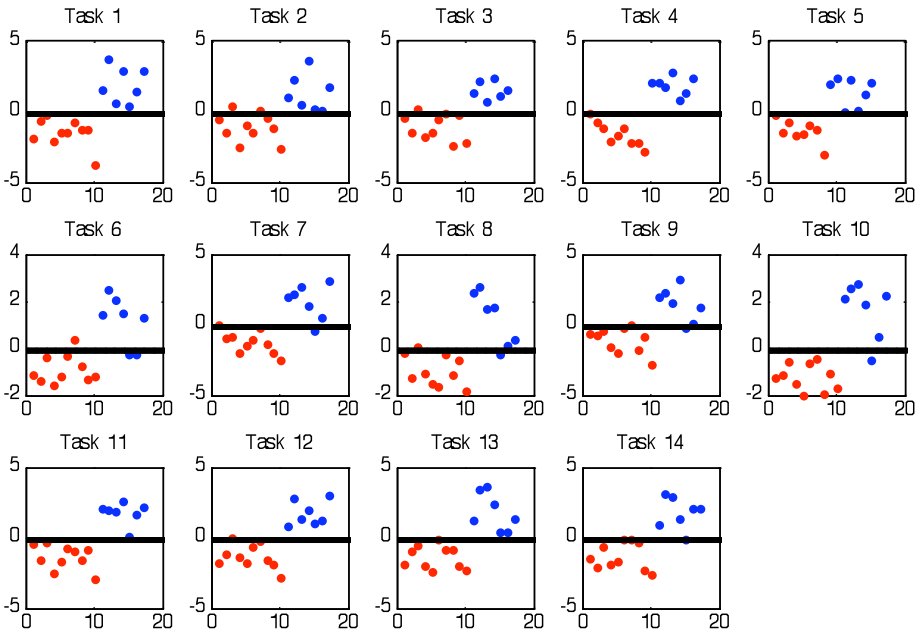
Bayesian Elastic Net (Modified label)
(\lambda_{2}=1)

Gene identified (4):

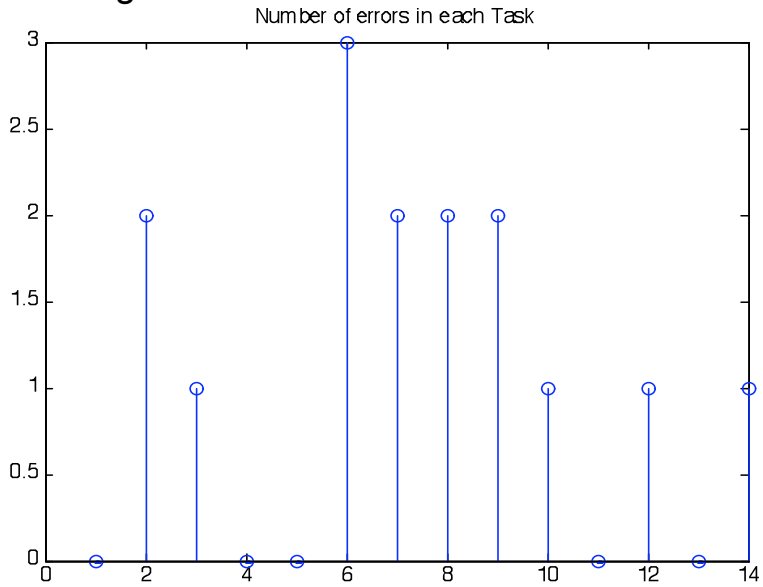
'ENOSF1'	'55556_at'
'HLA-DOB'	'3112_at'
'C13orf18'	'80183_at'
'LILRB2'	'10288_at'



Regression coefficients for each task



Regression response for each task

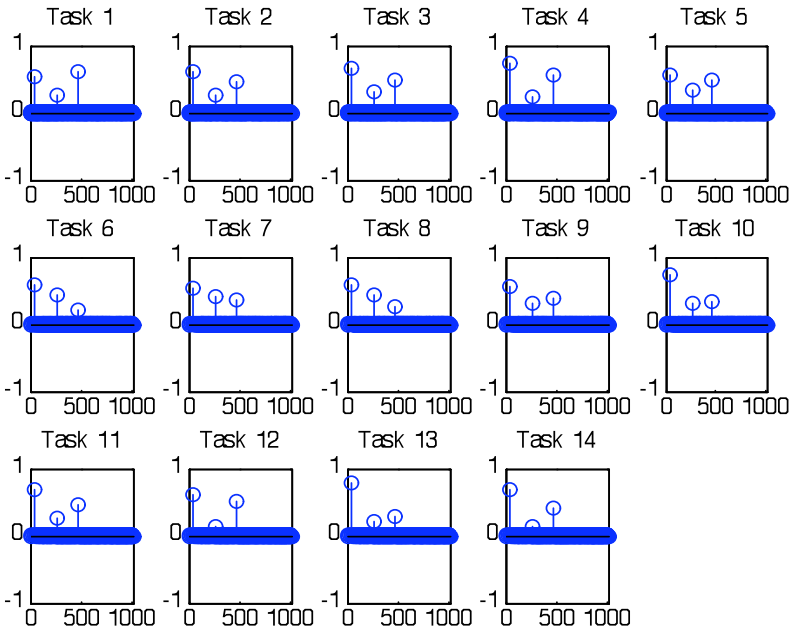


Number of errors in each task

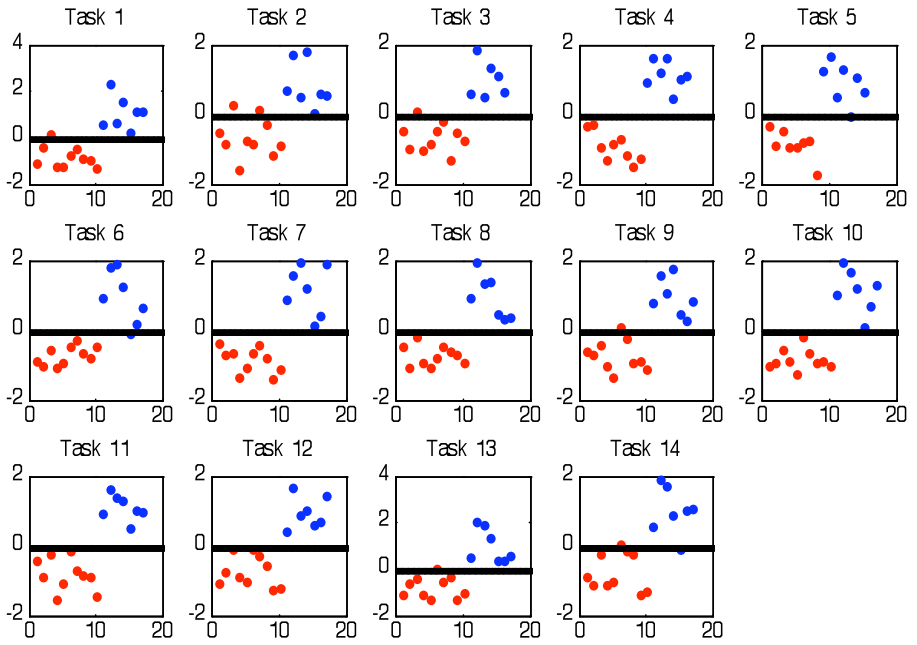
Bayesian LASSO (Original label)

Gene identified (2):

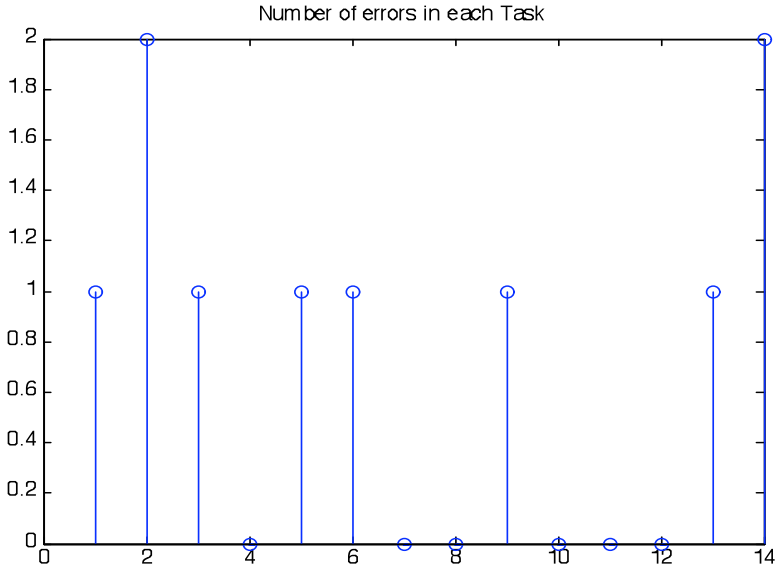
'HRASLS3' '11145_at'
'FADS2' '9415_at'



Regression coefficients for each task



Regression response for each task



Number of errors in each task

Bayesian Elastic Net (Original label)
(\lambda_{2}=1)

Gene identified (3):

'HRASLS3' '11145_at'
'FADS2' '9415_at'
'GM2A' '2760_at'