HW7

Shu Bin (I checked the sol'ns)

1) a) Let i be the datapoint index, k be the component, j be the dimension index of the D dimensional bit vector.

Then the log ~~likihood~~ likelihood:

$$\cancel{l(\mu)}\ l(\mu) = \sum_i \sum_k r_{ik} \log \mathbb{P}(x_i \mid \theta_k)$$
$$= \sum_i \sum_k r_{ik} \sum_j x_{ij} \log \mu_{kj} + (1 - x_{ij}) \log (1-\mu_{kj})$$

$$\frac{\partial l}{\partial \mu_{kj}} = \sum_i r_{ik} \left(\frac{x_{ij}}{\mu_{kj}} - \frac{1-x_{ij}}{1-\mu_{kj}} \right) = \frac{1}{\mu_{kj}(1-\mu_{kj})} \sum_i r_{ik} (x_{ij} - \mu_{kj}) = 0$$

thus the optimality condition is $\sum_i r_{ik} x_{ij} = \mu_{kj} \sum_i r_{ik}$

b) $l(\mu) = \sum_i \sum_k r_{ik} \log \mathbb{P}(x_i \mid \mu_k) + \log \mathbb{P}(\mu_k)$

$$= \sum_i \sum_k r_{ik} \left(\sum_j x_{ij} \log \mu_{kj} + (1 - x_{ij}) \log (1-\mu_{kj}) \right) + (a-1) \log \mu_{kj} + (b-1) \log (1-\mu_{kj})$$

$$\frac{\partial l}{\partial \mu} = \sum_i \left(\frac{r_{ik} x_{ij} + a - 1}{\mu_{kj}} - \frac{r_{ik}(1-x_{ij}) + b - 1}{1-\mu_{kj}} \right)$$

$$= \frac{1}{\mu_{kj}(1-\mu_{kj})} \left[\sum_i r_{ik} x_{ij} - \left(\sum_i r_{ik} + a + b - 2 \right) \mu_{kj} + a - 1 \right] = 0$$

thus the optimality condition is $\sum_i r_{ik} x_{ij} + a - 1 = \left(\sum_i r_{ik} + a + b - 2 \right) \mu_{kj}$

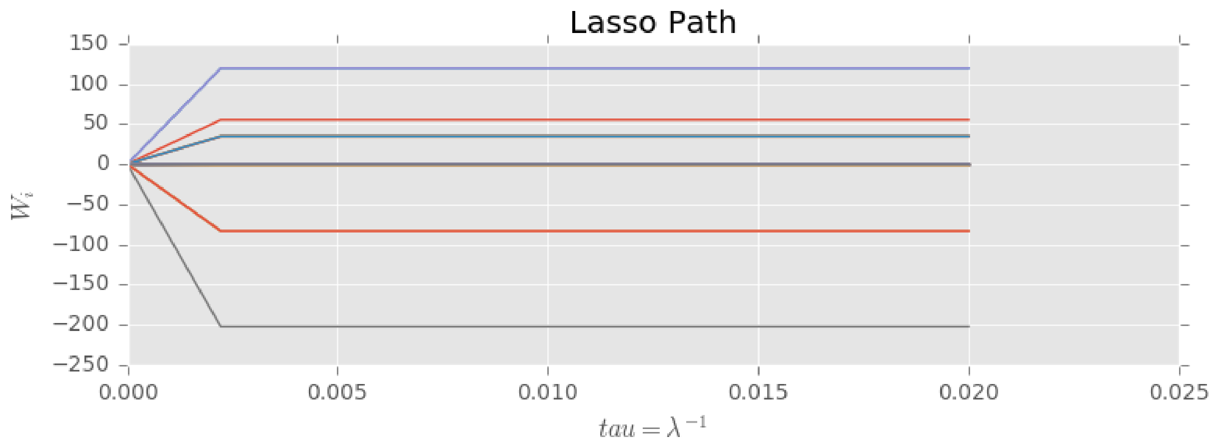
2) Let γ be the learning rate, then

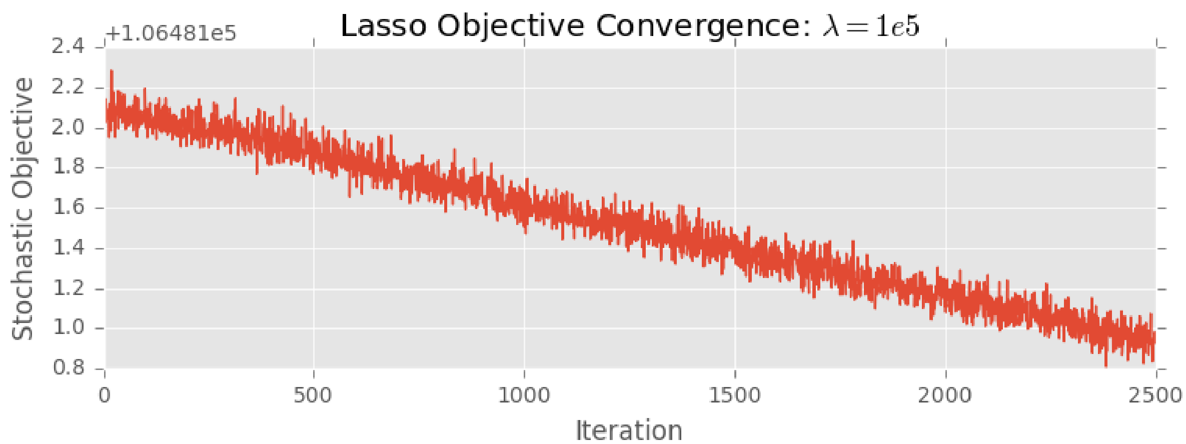
$$\mathrm{prox}_\gamma(x)_i = \begin{cases} x_i - \gamma & x_i > \gamma \\ 0 & |x_i| \le \gamma \\ x_i + \gamma & x_i < -\gamma \end{cases}$$

then $\frac{\partial \|x\|_1}{\partial x_i} = \frac{\partial \sum |x_i|}{\partial x_i} = \mathrm{sign}(x_i) \Rightarrow \nabla \|x\|_1 = \mathrm{sign}(x_i)$

then $\nabla \|Ax - b\|_2^2 + \lambda \|x\|_1 = \nabla x^T A^T A x - 2b^T A x + b^T b + \lambda \|x\|_1$

$$= 2A^T A x - 2b^T A + \lambda\, \mathrm{sign}(x)$$


Lasso Path
150
100
50
0
−50
−100
−150
−200
−250
W_i
0.000
0.005
0.010
0.015
0.020
0.025
$tau = \lambda^{-1}$


Lasso Objective Convergence: $\lambda = 1e5$
+1.06481e5
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
Stochastic Objective
0
500
1000
1500
2000
2500
Iteration