HW #7

1.

(a) Complete data log likelihood:

$$\ell(\mu) = \sum_i \sum_k r_{ik} \log \mathbb{P}(x_i \mid \theta_k)$$

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

$i \rightarrow$ datapoint index

$j \rightarrow$ dimension index of D dimensional bit vectors

$k \rightarrow$ component

Differentiating wrt μ_{kj}

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

$$\vdots$$

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

$$\sum_i r_{ik} x_{ij} = \mu_{kj} \sum_i r_{ik}$$

QED

(b) Complete data log + log prior:

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

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

$$(a-1) \log(\mu_{kj}) + (b-1) \log(1 - \mu_{kj})$$

Differentiating wrt μ

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

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

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

QED

2. $\nabla \|x\|_1 = \nabla \sum_i |x_i| = \sum_i \text{Sign}(x_i)\hat{e} = \text{Sign}(x)$

Taking gradient & setting to 0

$$\nabla\left(\|Ax-b\|_2^2 + \lambda\|x\|_1\right) = 0 = \lambda\,\text{Sign}(x) + 2\left(A^TAx - b^TA\right)$$

$$\Rightarrow \frac{\lambda}{2}\,\text{Sign}(x) = A^TAx - b^TA$$

The top 5 features are:

'timedelta'
'weekday-is-wednesday'
'weekday-is-thursday'
'weekday-is-friday'
'weekday-is-saturday'

from running the code.

(See attached convergence & regularization plots)