Hw5

2018年11月12日 星期一　　下午7:18

Question 1.

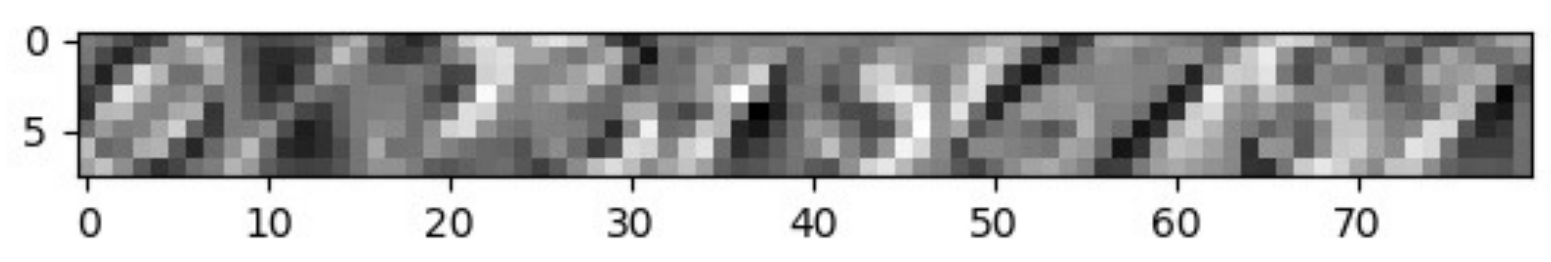
(a) trainning set average conditional log-likelihood is -0.1246244

testing set average conditional log-likelihood is -0.1966732

(b) The accuracy for train set is 0.98142

The accuracy for test set is 0.97275

(c)



Question 2)

(a) Assume we have up to n datapoints and K categorial

$$P(\theta|D) \propto P(D|\theta) P(\theta)$$

$$\propto \prod_{i=1}^{n} \prod_{j=1}^{K} \theta_j^{\mathbb{1}(x_{ij}=1)} \prod_{j=1}^{K} \theta_j^{\alpha_j - 1}$$

$$= \prod_{j=1}^{K} \theta_j^{N_j} \prod_{j=1}^{K} \theta_j^{\alpha_j - 1}$$

$$= \prod_{j=1}^{K} \theta_j^{(N_j+\alpha_j)-1}$$

where $N_j = \sum_{i=1}^{n} \mathbb{1}(x_{ij}=1)$

(b) MAP is trying to maximize the posterior

Therefore. $\hat{\theta}_{MAP} = \arg\max(P(\theta|D))$

$$\hat{\theta}_{MAP} = \arg\max\left(\prod_{j=1}^{K} \theta_j^{(N_j+\alpha_j)-1}\right)$$

Applying logarithm, we have

$$\hat{\theta}_{MAP} = \arg\max\left(\log\left(\prod_{j=1}^{K} \theta_j^{(N_j+\alpha_j)-1}\right)\right)$$

$$= \arg\max\left(\sum_{j=1}^{K} (N_j+\alpha_j-1) \log(\theta_j)\right)$$

We know that $\sum_{j=1}^{K} \theta = 1$, therefore we have the constrain of $\sum_{j=1}^{K} \theta_j - 1 = 0$.

Applying Lagrange multiplier, we have

$$L(\theta,\lambda) = \sum_{j=1}^{K} (N_j+\alpha_j-1) \log(\theta_j) - \lambda\left[\sum_{j=1}^{K} \theta_j - 1\right]$$

$$\frac{\partial L(\theta,\lambda)}{\partial \theta_j} = \frac{(N_j+\alpha_j-1)}{\theta_j} - \lambda$$

Setting $\frac{\partial L(\theta,\lambda)}{\partial \theta_j} = 0$, we have

$$\lambda = \frac{N_j+\alpha_j-1}{\theta_j}$$

Therefore, for each θ, we have $\theta_j = \frac{N_j+\alpha_j-1}{\lambda}$

put it back to constrain that $\sum_{j=1}^{K} \theta_j - 1 = 0$.

we have $\sum_{j=1}^{K} \frac{N_j+\alpha_j-1}{\lambda} - 1 = 0$.

$$\therefore \quad \lambda = \sum_{j=1}^{K} N_j + \alpha_j - 1$$

Finally, substitude λ back to θ which we derived.

$$\hat{\theta}_{MAP} = \frac{N_j+\alpha_j-1}{\sum_{j=1}^{K} (N_j+\alpha_j-1)}$$