CSC 411 H1 FALL 2018

LINO LASTELLA 1001237654

HW 5

2 (d)

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

$$\propto \theta_1^{a_1-1} \cdot \ldots \cdot \theta_K^{a_K-1} \cdot \prod_{i=1}^{N} \prod_{k=1}^{K} \theta_k^{x_k^{(i)}}$$, because data is i.i.d.

$$= \theta_1^{a_1-1} \cdots \theta_K^{a_K-1} \cdot \theta_1^{x_1^{(1)}} \cdots \theta_K^{x_K^{(1)}} \cdots \theta_1^{x_1^{(N)}} \cdots \theta_K^{x_K^{(N)}}$$

$$= \prod_{k=1}^{K} \theta_k^{a_k - 1 + \sum_{i=1}^{N} x_k^{(i)}}$$

This is another Dirichlet distribution with parameters

$$\alpha_i = a_i + \sum_j x_i^{(j)} .$$

$\theta_{pred} = \Pr(x' = k \mid D)$, where x' is the next outcome

$$= \int p(\theta_k|D) \Pr(x' = k \mid \theta) \, d\theta_k$$

$$= \int \text{Dirichlet}(\theta_k; a_1 + \sum_j x_1^{(j)}, \ldots, a_K + \sum_j x_K^{(j)}) \cdot \theta_k \, d\theta_k$$

$= \mathbb{E}[\theta_k]$, where $\mathbb{E}[\theta_k] = \mathbb{E}_{\text{Dirichlet}(\theta_k; a_1 + \sum_j x_1^{(j)}, \ldots, a_K + \sum_j x_K^{(j)})}[\theta_k]$

$$= \frac{a_k + \sum_j x_k^{(j)}}{\sum_{k'} \left(a_{k'} + \sum_j x_{k'}^{(j)}\right)}$$

(b)

$$\hat{\theta}_{MAP} = \arg\max_{\theta} p(\theta|D)$$

$$= \arg\max_{\theta} p(\theta, D)$$

$$= \arg\max_{\theta} p(\theta)\, p(D|\theta)$$

$$= \arg\max_{\theta} \left(\log p(\theta) + \log p(D|\theta)\right)$$

$$= \arg\max_{\theta} \left[\text{const} + \sum_{k=1}^{K} (\alpha_k - 1) \log \theta_k + \sum_{i=1}^{N} \sum_{k=1}^{K} x_k^{(i)} \log \theta_k \right]$$

$$= \arg\max_{\theta} \left[\text{const} + \sum_{k=1}^{K} \left(\alpha_k - 1 + \sum_{i=1}^{N} x_k^{(i)} \right) \log \theta_k \right]$$

We can maximize by applying Lagrange Multiplier Theorem

$$\frac{d}{d\theta_k} \left[\sum_{k=1}^{K} \left(\alpha_k - 1 + \sum_{j} x_k^{(j)} \right) \log \theta_k \right] + \frac{d}{d\theta_k} \left[-\lambda \cdot \sum_{i} (\theta_i - 1) \right] = 0$$

$$\rightarrow \frac{\alpha_k - 1 + \sum_{j} x_k^{(j)}}{\theta_k} - \lambda = 0 \rightarrow \theta_k = \frac{\alpha_k - 1 + \sum_{j} x_k^{(j)}}{\lambda} .$$

Since $\sum_{k} \theta_k = 1$, $\lambda = \sum_{k'=1}^{K} \left(\alpha_{k'} - 1 + \sum_{j} x_{k'}^{(j)} \right)$

$\therefore$ $\hat{\theta}_{MAP} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_K \end{pmatrix}$ such that

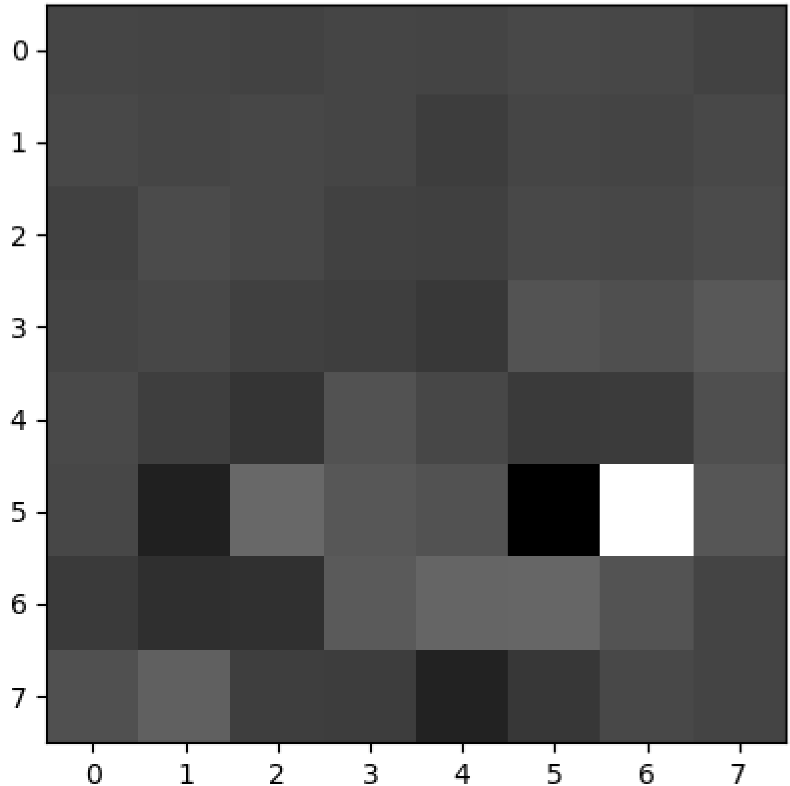
$$\theta_i = \frac{\alpha_i - 1 + \sum_{j} x_i^{(j)}}{\sum_{k=1}^{K} \left(\alpha_k - 1 + \sum_{j} x_k^{(j)} \right)} , \quad \forall i = 1, 2, \cdots, K .$$

$\square$

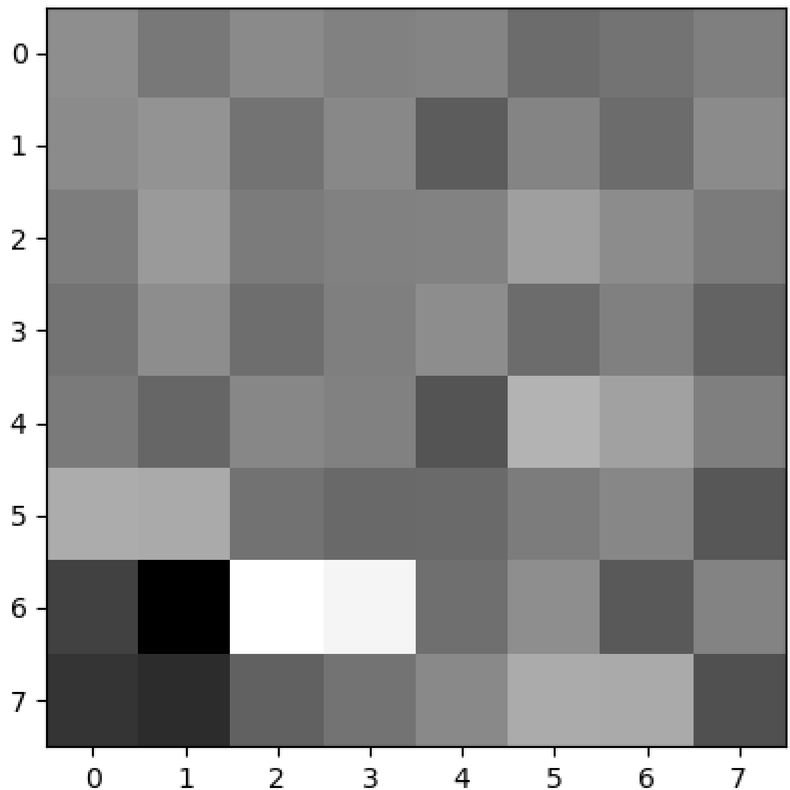
Question 1

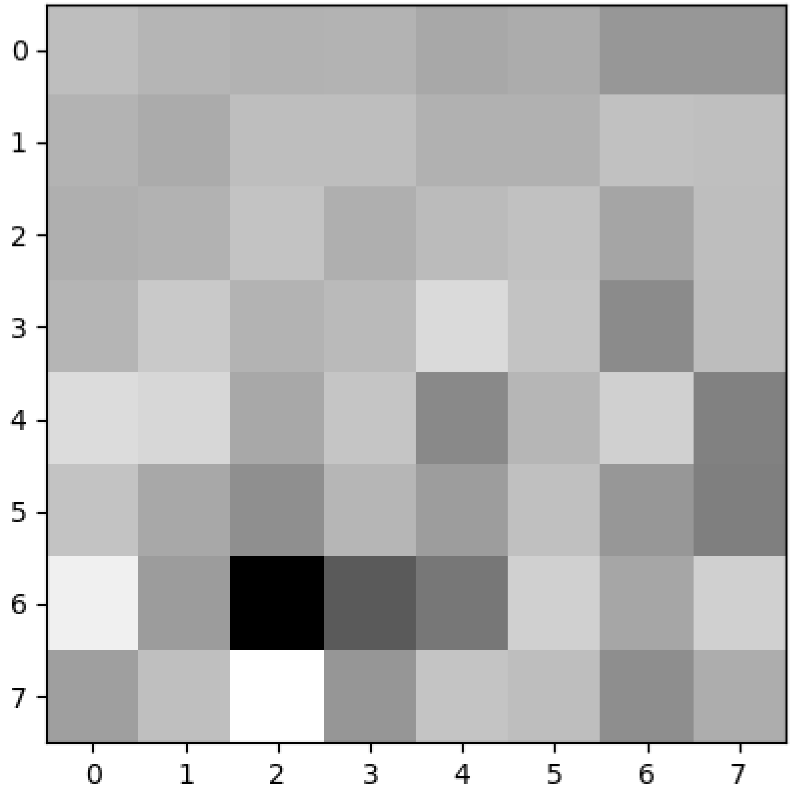
A)
Training set ---> -0.12462443666863024
Test set ---> -0.19667320325525567

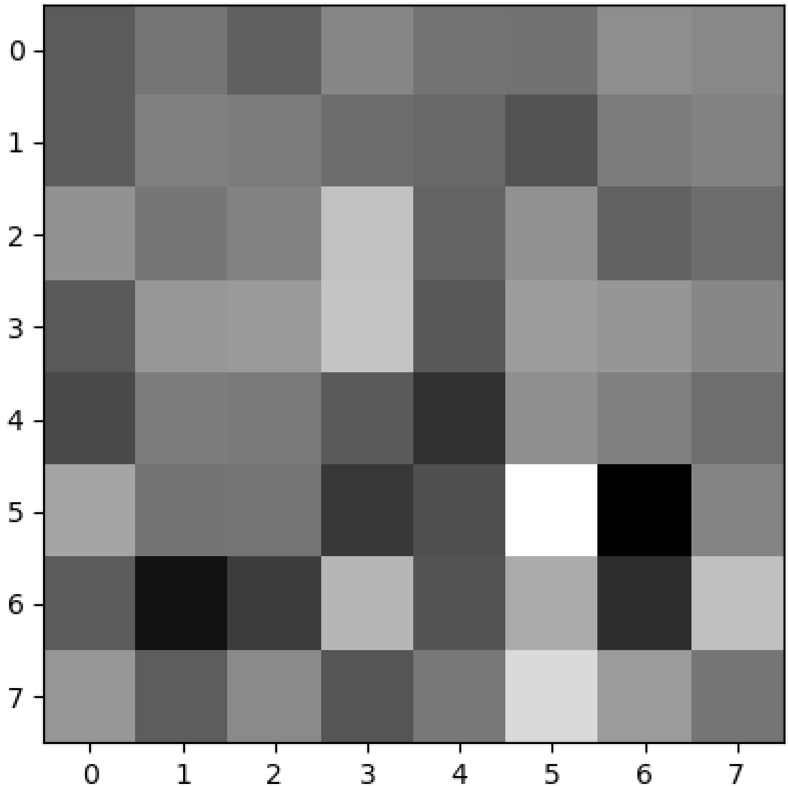
B)
Training set ---> 98.1428571429
Test set ---> 97.275

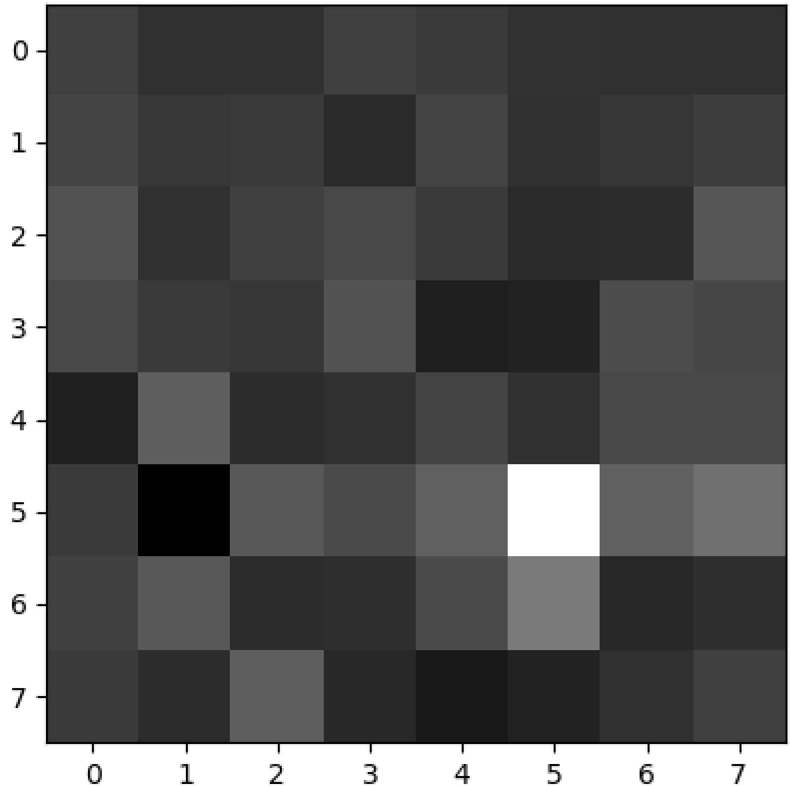


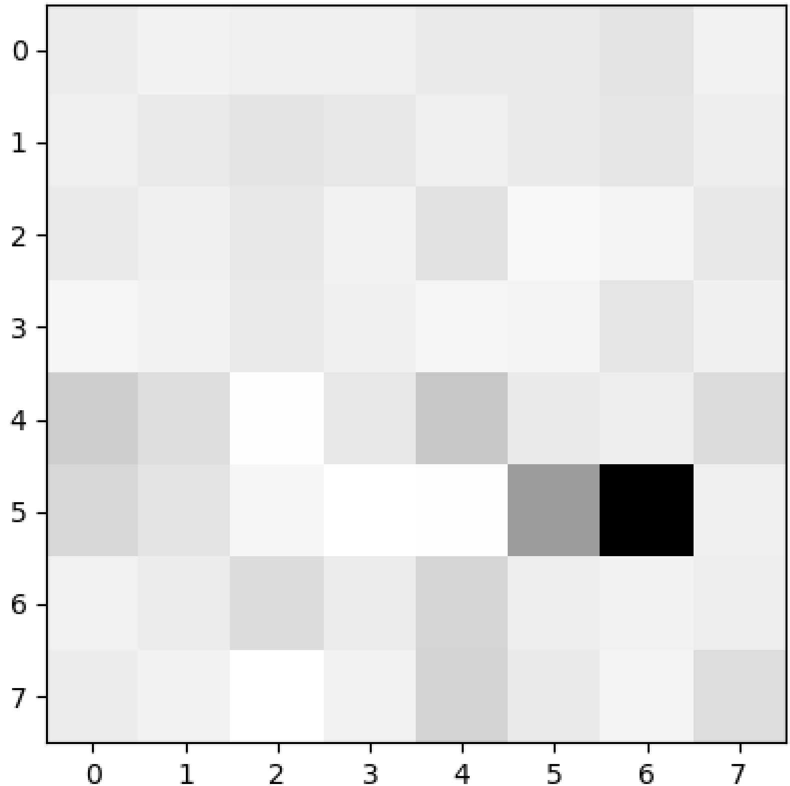
0
1
2
3
4
5
6
7
0
1
2
3
4
5
6
7

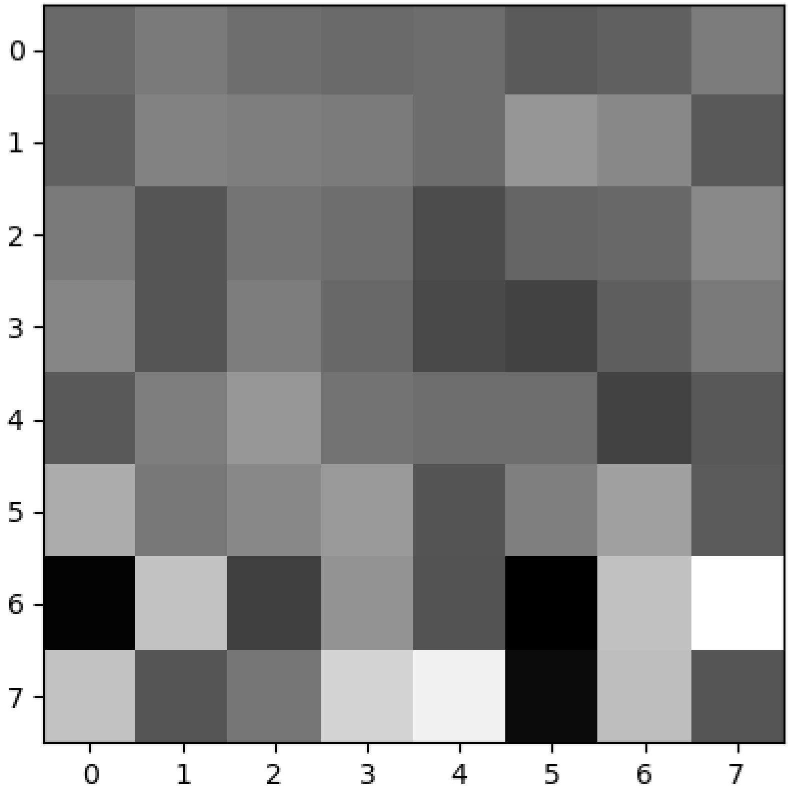

0
1
2
3
4
5
6
7
0
1
2
3
4
5
6
7

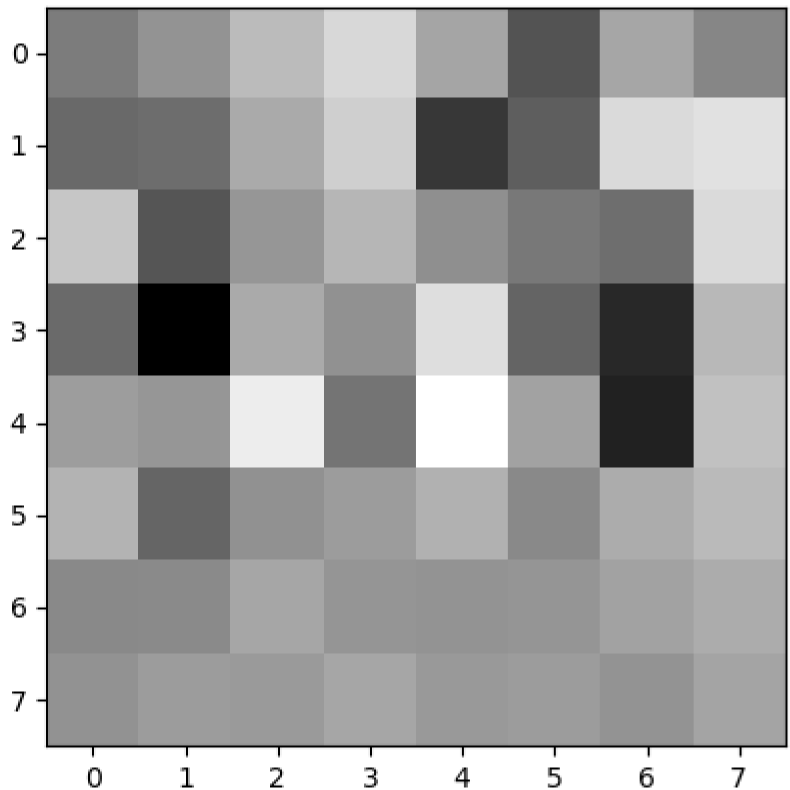

0
1
2
3
4
5
6
7
0
1
2
3
4
5
6
7


0
1
2
3
4
5
6
7
0
1
2
3
4
5
6
7

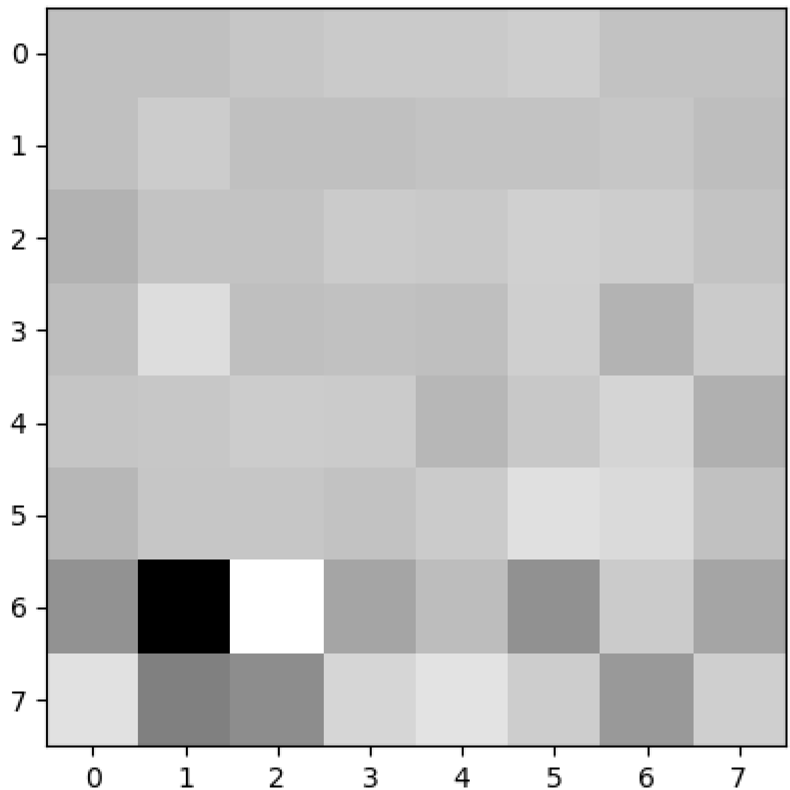

0
1
2
3
4
5
6
7
0
1
2
3
4
5
6
7

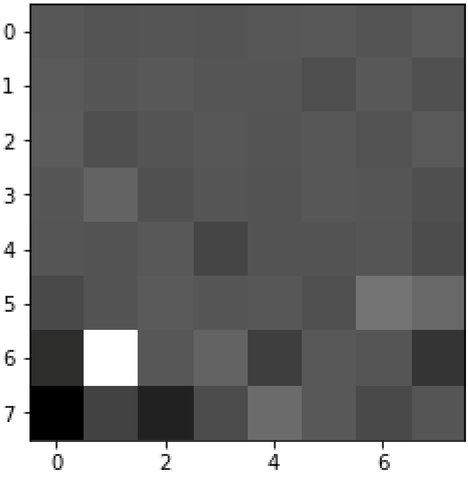

0
1
2
3
4
5
6
7
0
1
2
3
4
5
6
7


0
1
2
3
4
5
6
7
0
1
2
3
4
5
6
7



0
1
2
3
4
5
6
7
0
1
2
3
4
5
6
7


0
1
2
3
4
5
6
7


0
1
2
3
4
5
6
7
0
2
4
6