

MLE for the One-Hot Categorical

$$\underline{\Theta}^* = \underset{\substack{\Theta \in [0,1]^D \\ \sum_{d=0}^{D-1} \theta_d = 1}}{\operatorname{argmax}} (\ell(\underline{D}; \underline{\Theta}))$$

$$\Rightarrow \theta_e^* = \frac{N_e}{N}$$

Dataset for the one-hot

$$\underline{D} = \begin{Bmatrix} [0, 0, 1], \\ [0, 1, 0], \\ [0, 0, 1], \\ [1, 0, 0], \\ \vdots \end{Bmatrix}$$

View the dataset as a matrix

$$\underline{D} = \begin{array}{|c|c|c|} \hline 0 & 0 & 1 \\ \hline 1 & 0 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 0 & 1 \\ \hline \end{array} \left. \vphantom{\begin{array}{|c|c|c|}} \right\} \begin{array}{l} \text{\# of samples } (N) \\ \text{\# of states } (D) \end{array}$$

reduce matrix to vector over
0th axis

$$\hookrightarrow \underline{N} = [N_0, N_1, N_2, \dots]$$

$$\underline{\Theta} = \left[\frac{N_0}{N}, \frac{N_1}{N}, \frac{N_2}{N}, \dots \right]$$