



Why

1

Definition

Let Z and X be non-empty finite sets. We want to model a distribution $p^\theta : Z \times X \rightarrow \mathbf{R}$. We parameterize a family of distributions by a parameter θ . We have a dataset (x^1, \dots, x^n) . Given a parameter θ^0 , we want to solve

$$\begin{array}{ll} \text{find} & \theta \\ \text{to maximize} & \sum_{k=1}^n \mathbf{E}_{p_{z|x}^{\theta^0}(z, x^k)} [\log p^\theta(z, x)] \end{array}$$

Binary Gaussian Mixture Example

2

¹Future editions will rework this sheet.

²Future editions will expand.

