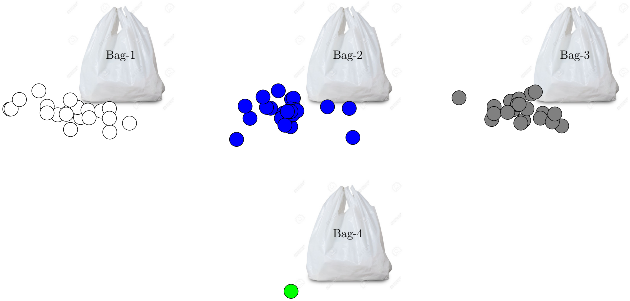
IIT Jodhpur

Biological Vision and Applications
Module 03-09: Hierarchical Bayesian Model

Hiranmay Ghosh

An example

The bags can have marbles



Specific knowledge and Generic knowledge

- Specific Knowledge
  - When we sample marbles from a particular bag, we gain knowledge about that bag
  - e.g. uniformity of colors and color of marbles in that bag
- Generic Knowledge
  - When we sample marbles from several bags, we gain knowledge about all bags
    - ... even those are not sampled
  - e.g. uniform color of marbles in each bag
- Specific knowledge about several bags lead to generic knowledge
  - This is an instance of inductive reasoning or inductive generalization
  - The process of gaining generic knowledge is also known as meta-learning

Modeling the problem

- Let $\vec{\theta}_i$ represent the model parameters for bag i
  - $\vec{\theta}_i = (\theta_{ij}, j = 1 \ldots n)$
  - j represents the different colors. $\sum_j \theta_{ij} = 1$
- In HBM
  - $\vec{\theta}_i$s are modeled as probabilistic functions of some hyper-parameters
  - The hyper-parameters represent a higher (more abstract) level of knowledge
- A common approach is to use Dirichlet distribution
  - In this example, parameters can be $\alpha, \vec{\beta}$
  - α represents the heterogeneity of colors of the marbles in the individual bags
  - $\vec{\beta}$ representing the average color distribution across all the bags

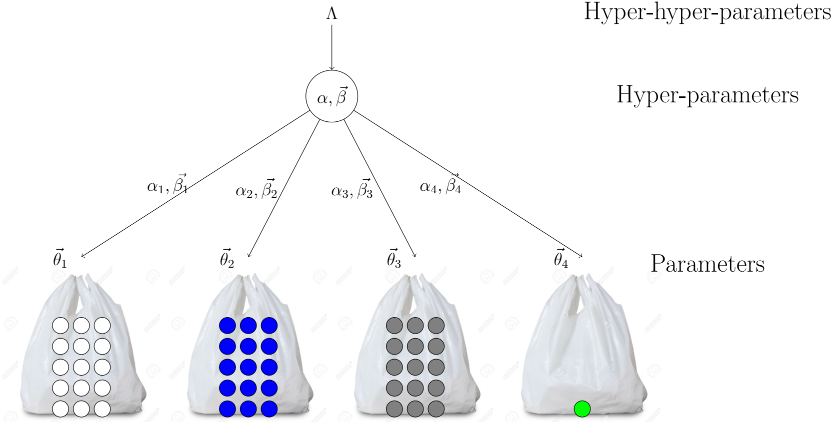
On Dirichlet distribution

- Beta Distribution
  - A probability distribution function with two parameters (α, β)
  - $p(\theta)_{\alpha,\beta} = \frac{1}{k}.\theta^{\alpha-1}.(1-\theta)^{\beta-1}$
  - where $k = \frac{\Gamma(\alpha).\Gamma(\beta)}{\Gamma(\alpha+\beta)}$
  - where $\Gamma(x) = \int_{t=0}^{\infty} t^{x-1}e^{-t}dt$
- Dirichlet Distribution (a generalization of Beta distribution)
  - $\vec{\theta} = (\theta_1, \theta_2 \ldots \theta_n)$
  - $\vec{\alpha} = (\alpha_1, \alpha_2 \ldots \alpha_n)$
  - $p(\vec{\theta})_{\vec{\alpha}} = \frac{1}{k}.\prod_{i=1}^{n} \theta_i^{\alpha_i-1}$
  - where $k = \frac{\prod_{i=1}^{n}\Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{n}\alpha_i)}$

https://leimao.github.io/blog/Introduction-to-Dirichlet-Distribution/

Hierarchical Bayesian Model

A graphical depiction



Discussions

- The models for the bags are linked with hyper-parameters $\alpha, \vec{\beta}$
  - Are learned together with θs
  - An observation for one bag serves as an observation for the other bags too
  - Hyper-parameters $\alpha, \vec{\beta}$ are learned together with the parameters θ_is

- $\vec{\theta}_i$ is a probabilistic function of $\alpha, \vec{\beta}$
  - $\alpha, \vec{beta}$ impose constraints on values of $\vec{\theta}_i$s
  - Priors for $\vec{\theta}_i$s (no observations) are closer to actual values
  - $\vec{\theta}_i$ can be learned (reliably estimated) from less number of observations

- Hyper-parameters represent more abstract knowledge
- It is possible to model $\alpha, \vec{\beta}$ with even higher level of knowledge ...
  - Further inductive generalization is possible
  - Generalization from one problem to another will be efficient for similar problems

Quiz

Quiz 03-09

End of Module 03-09