

## 0.1 K-means

## 0.2 Hierarchical clustering

## 0.3 Nonparametric clustering

### 0.3.1 Dirichlet processes

- DD is a distribution of distributions, each sample from DD is a categorical distribution over K categories
- DD is parameterized by  $G_0$  and  $\alpha$  a scale factor
- When  $\alpha$  is large, samples from  $DD(\alpha \cdot G_0)$  will be close to  $G_0$
- Dirichlet process is used to cluster data *without specifying the number of clusters in advance*
- nonparametric because its dimensionality is infinite
- exhibits a rich-gets-richer property
- observations are probabilistically assigned to clusters based on the # of observations in that cluster

$$P(\text{cluster} = k) = \frac{n_k}{\alpha + n - 1}$$

$$P(\text{cluster} = \text{new}) = \frac{\alpha}{\alpha + n - 1}$$

- [DP tutorial](#)
- [Another DP tutorial](#)

### 0.3.2 Chinese restaurant process

- Restaurant starts off empty
- First person selects a group
- Second person sits at a new table with probability  $\frac{\alpha}{\alpha+1}$  and sits with the first person with probability  $\frac{1}{\alpha+1}$

### 0.3.3 Polya Urn Model

- Same model as CRP
- urn contains  $\alpha G_0$  balls of color x for each x
- at each timestep, draw a ball from the urn and drop it back into the urn plus another ball of the same color
- CRP specifies only a distribution over partitions, but does not assign parameters to each group whereas the Polya Urn Model does both

### 0.3.4 Stick-breaking construction

- Figure out the proportion of points that fall into a particular group
- Start with a stick of length 1
- Generate a random variable  $\beta_1 \sim \text{Beta}(1, \alpha)$  and break off the stick at  $\beta_1$
- Take the stick to the right and repeat
- Stick-breaking is CRP or Polya Urn from a different perspective

### 0.3.5 Gibbs sampling

### 0.3.6 Metropolis Hastings

- Bayesian clustering algorithms often rely on the Dirichlet Distribution (DD) to encode prior information about cluster assignments