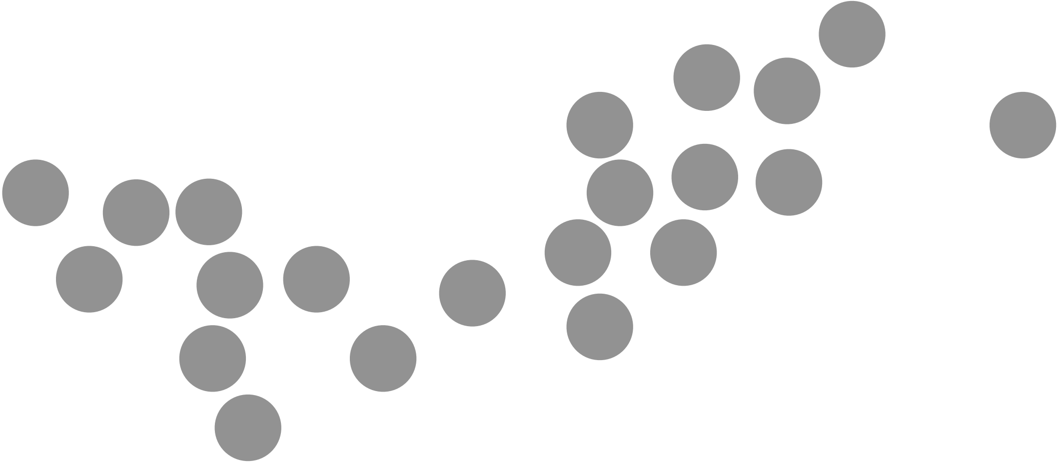
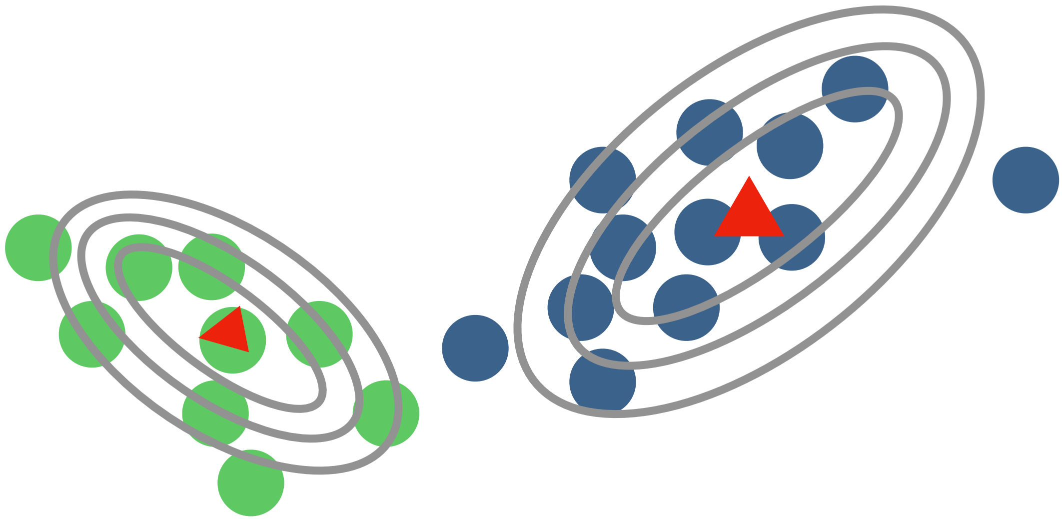
Background: Bayesian inference

2



have some data

$$y_n \sim w_1 \mathcal{N}(\mu_1, \Sigma_1) + w_2 \mathcal{N}(\mu_2, \Sigma_2)$$



wish to infer parameters and z_n, the cluster label for each datum:

$$x = (w_1, w_2, \mu_1, \mu_2, \Sigma_1, \Sigma_2, (z_n))$$

Bayesian inference: compute conditional distribution of parameters given data

posterior distribution
(beliefs after observing data)

$$p(x \mid y)$$

$$p(x \mid y) \propto p(y \mid x)p(x)$$

likelihood

(GMM)

prior

(beliefs before data)

problem: $p(x \mid y)$ intractable

solution: approximate it with some $q(x)$ *(variational inference)*

- lots of work if x is continuous (e.g., normalizing flows)

- little work for discrete x

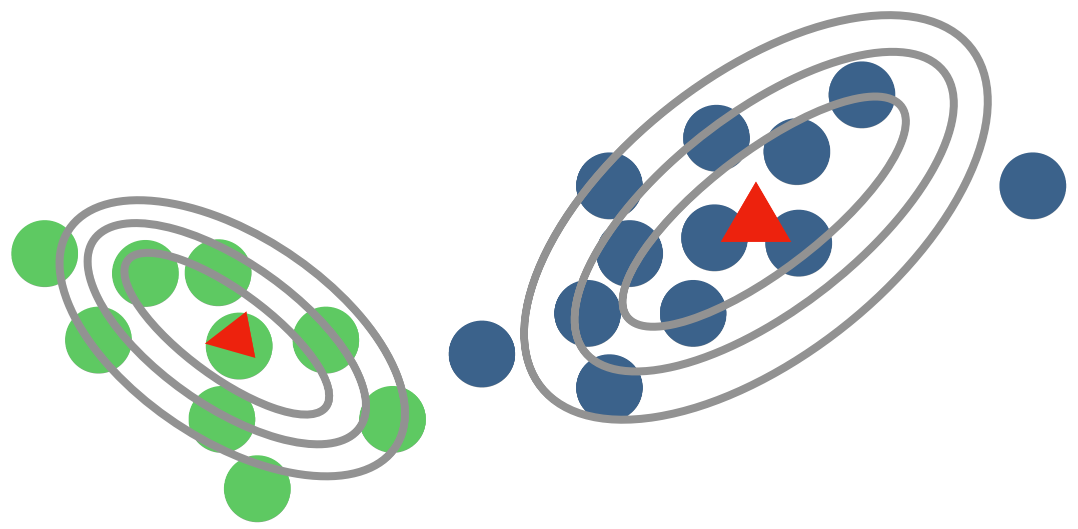
usually, embed x in a continuous space and use known methods

this talk: variational inference for discrete posteriors without continuous embedding

Background: Bayesian inference

have some data

$y_n \sim w_1 \mathcal{N}(\mu_1, \Sigma_1) + w_2 \mathcal{N}(\mu_2, \Sigma_2)$



wish to infer parameters and z_n, the cluster label for each datum:

$x = (w_1, w_2, \mu_1, \mu_2, \Sigma_1, \Sigma_2, (z_n))$

Bayesian inference: compute conditional distribution of parameters given data

posterior distribution
(beliefs after observing data)

$$p(x \mid y) \propto p(y \mid x) p(x)$$

likelihood
(GMM)

prior
(beliefs before data)

problem: $p(x \mid y)$ intractable

solution: approximate it with some $q(x)$ (*variational inference*)

- lots of work if x is continuous (e.g., normalizing flows)
- little work for discrete x

usually, embed x in a continuous space and use known methods

this talk: variational inference for discrete posteriors without continuous embedding

Background: normalizing flows