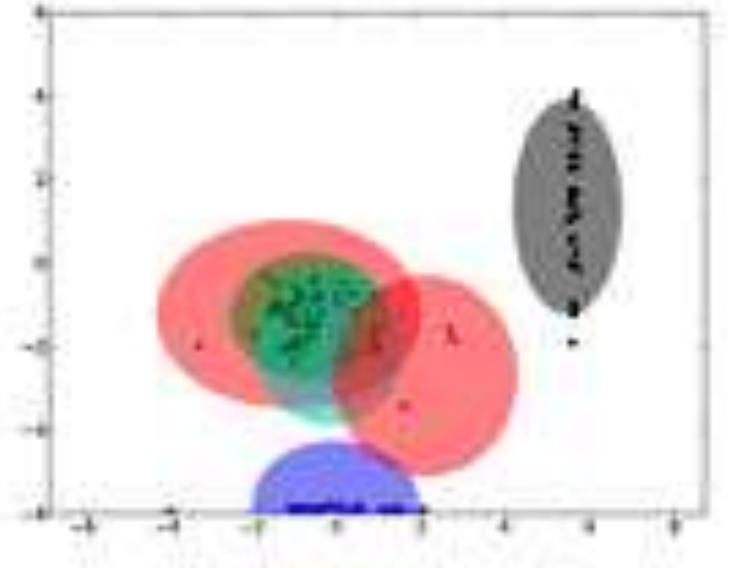
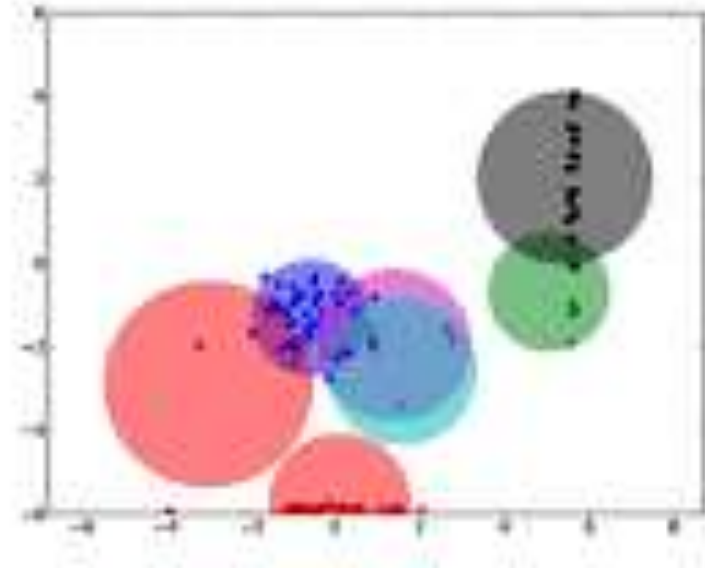
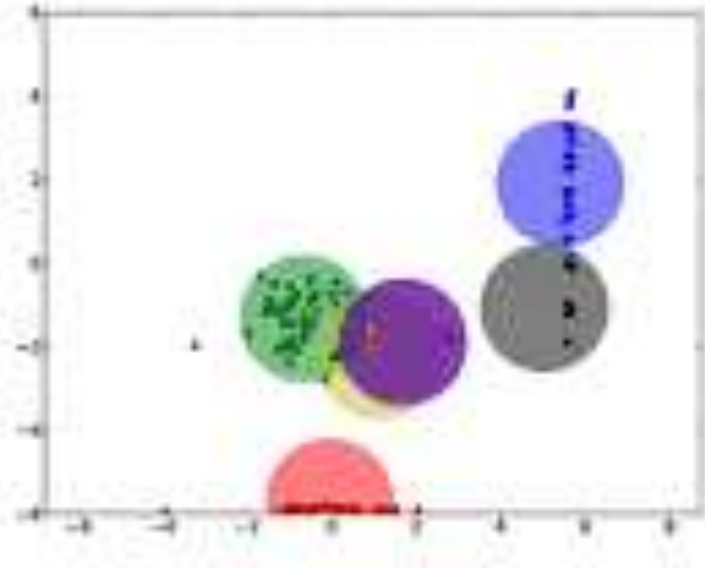


Dirichlet Process

- ❖ The Dirichlet process (DP) is a stochastic process used in Bayesian nonparametric models of data, particularly in Dirichlet process mixture models (also known as infinite mixture models).
- ❖ It is a distribution over distributions, that is, each draw from a Dirichlet process is itself a distribution
- ❖ It is called a Dirichlet process because it has Dirichlet distributed finite dimensional marginal distributions, just as the Gaussian process, another popular stochastic process used for Bayesian nonparametric regression, has Gaussian distributed finite dimensional marginal distributions.
- ❖ Distributions drawn from a Dirichlet process are discrete, but cannot be described using a finite number of parameters, thus the classification as a nonparametric model.

Dirichlet mixture model



Click here ↓

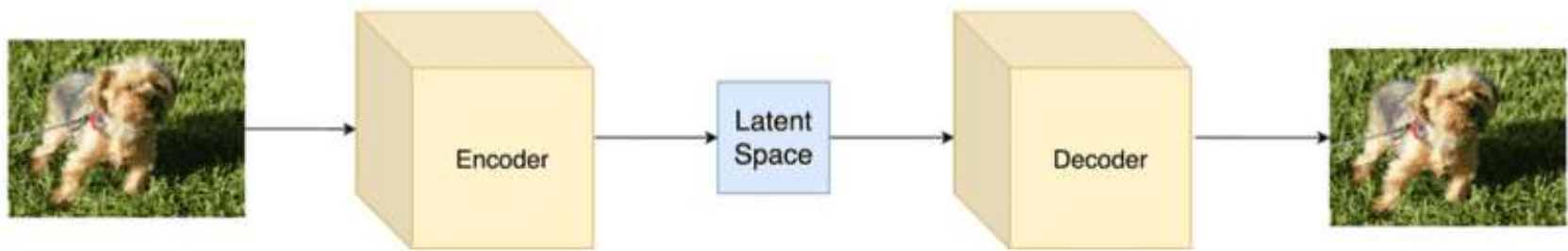
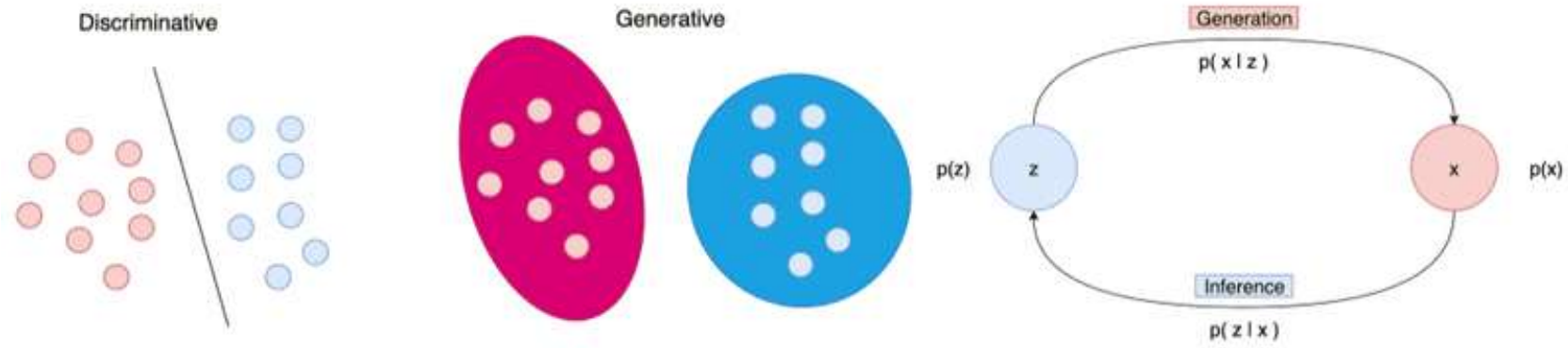
- <https://youtu.be/UTW530-QVxo>

Latent variable modeling

- ❖ A latent variable model, as the name suggests, is a statistical model that contains latent, that is, unobserved, variables.
- ❖ Their roots go back to Spearman's 1904 seminal work[1] on factor analysis, which is arguably the first well-articulated latent variable model to be widely used in psychology, mental health research, and allied disciplines
- ❖ A latent variable, defined in the broadest manner, is no more mysterious than an error term in a normal theory linear regression model or a random effect in a mixed model.

For example

- ❖ In a latent variable model for measuring level of depression (the latent variable of interest), the full range of clinician ratings or self-reported symptoms of mood disturbance, anhedonia, sleep disturbance, weight problems, psychomotor problems, worthlessness or guilt



Latent Dirichlet Allocation

- Latent Dirichlet Allocation (LDA) is a generative statistical model that explains a set of observations through unobserved groups, and each group explains why some parts of the data are similar.
- The LDA is an example of a topic model.
- In this, observations (e.g., words) are collected into documents, and each word's presence is attributable to one of the document's topics.
- LDA is an unsupervised learning algorithm that attempts to describe a set of observations as a mixture of different categories. These categories are themselves a probability distribution over the features.
- LDA is a generative probability model, which means it attempts to provide a model for the distribution of outputs and inputs based on latent variables. This is opposed to discriminative models, which attempt to learn how inputs map to outputs.

