



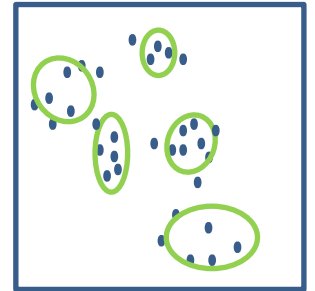
統計的機械学習:トピックモデル

ベイズ統計に基づく機械学習

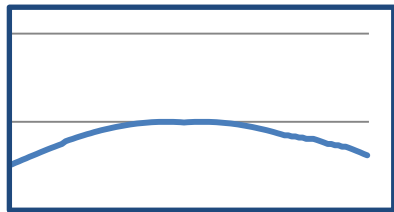


ノンパラメトリックベイズ

- K-means、EM、変分ベイズなどは、モデルの複雑さ、たとえばクラスタリングにおけるクラスタ数は予め分かっているとしてモデル推定した。しかし、実際はクラスタ数が不明の場合が多い。
- 観測データに応じてモデル自体の複雑さ(クラスタ数)も学習する
- 有限個の正規分布を配分比 π_k での混合モデル:
- クラスタ数 K の値を観測データから最適化することによって推定する。
- **基本的アイデア:無限次元の連続分布から観測データに適応した有限次元の離散分布を学習する。**



連続ドメインの
基底測度 G_0



連続ドメインの基底
測度を

- 可算無限次元の
離散分布 で近似し
て事前分布とする。



有限次元離散分
布

- 観測データか
ら推定

