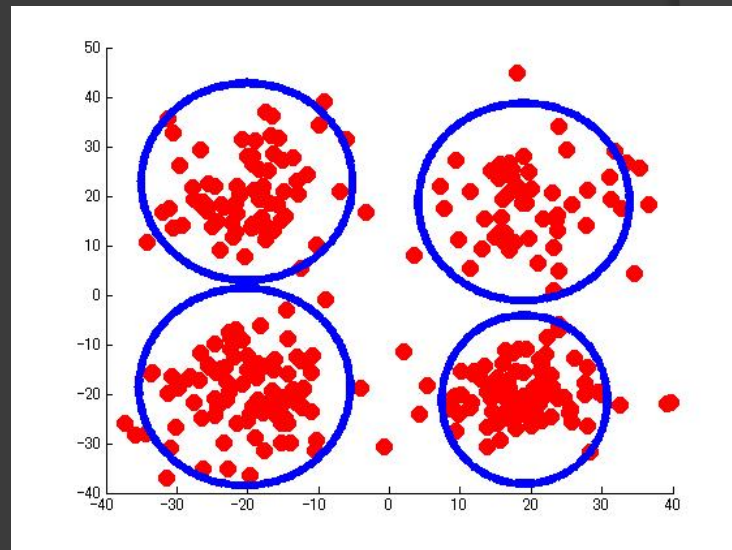
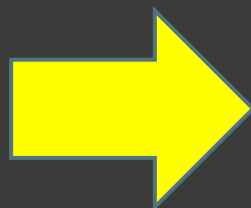
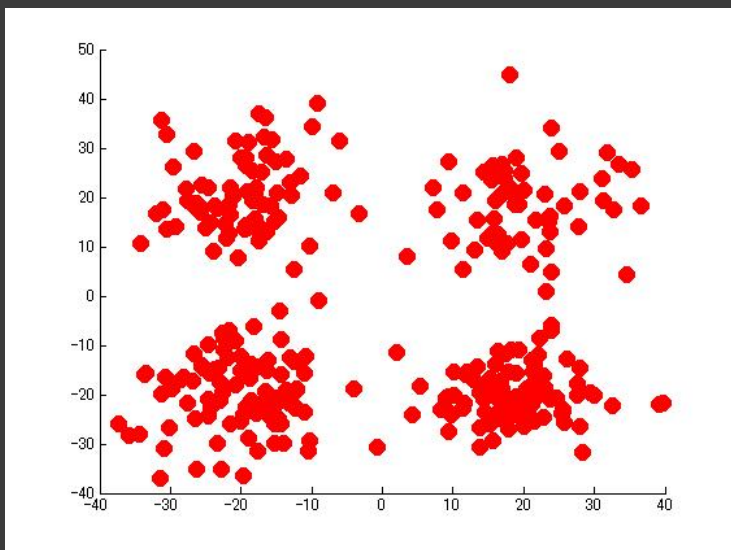


5-23 ガウス混合分布の正規化最尤符号の効率的計算法とモデル選択

平井 聡, 山西 健司 (東京大学)

モデル選択：データをガウス混合分布を用いてクラスタリング



最適なクラスタ数を知りたい



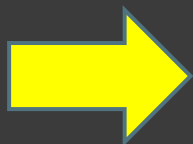
正規化最尤符号を用いる

正規化最尤(NML)符号：モデルに対するデータの最短記述長

⇒連続値データに対しては正規化項が無限に発散

$$P_{NML}(\mathbf{x}^n | \mathcal{M}) = \frac{P(\mathbf{x}^n | \hat{\theta}(\mathbf{x}^n, \mathcal{M}))}{\sum_{\mathbf{y}^n \in \mathcal{X}^n} P(\mathbf{y}^n | \hat{\theta}(\mathbf{y}^n, \mathcal{M}))}$$

発散



この問題を克服し、 $O(n^2)$ の時間で計算可能に



人工データでは、
データ数が少なくても
正解を導くことができた

ノード数100以下では、
BICと比べて平均約2.6倍のbenefit

