Subject /FH-Q2-6 Date /

Attendants

probablistic model

$p(x|\theta)$ where θ is vector

N traing examples x_i, $i \in [1, N]$

are drawn independently

from the

data distribution

6(a) Scheme of estimating θ

to maximize the log likelihood

of N traing

example.

生成モデルも識別モデルと同様に

$p(C_k|x)$ をモデル化することを考える

→ データに対するラベルの条件付き確率

$$p(C_k|x) = \frac{p(x|C_k)\,p(C_k)}{p(x)}$$

Subject 生成モデル Date 2

Attendants

$p(C_k|x) = \frac{p(x|C_k)\,p(C_k)}{p(x)}$

これが クラスに対する
データの条件付き分布になる

↓

クラスに応じた 擬似的なデータ をその
も作ることができる

識別モデル

$p(x|\theta)$

$$\hat{y} = \underset{y}{\operatorname{argmax}} P(y|x)$$

↑ これを直接モデル化

識別モデルは 直接確率を 直接求めるモデル

生成モデルでは $p(y|x)$ を直接はモデル化しない

事後分布 $p(y|x) =$

尤度 $P(x|y)$ 事前分布 $P(y)$

$p(x)$

この右辺が最大となる クラス y に分類する

分類時に

分母 $p(x)$ は分類には依存しないので

$$\hat{y} = \underset{y}{\operatorname{argmax}}\ p(x|y)p(y)$$

観測されたデータは無作為に生成されるのではなく

何らかの分布に基づいて生成

クラスの割合

全体のうちクラスAの データが60%なら

$p(y) = 0.6$

この分布が y のクラスごとに異なっており、各クラスがデータを生成する確率

Subject EMアルゴリズム（最尤推定の方法） Date 3-4

Attendants

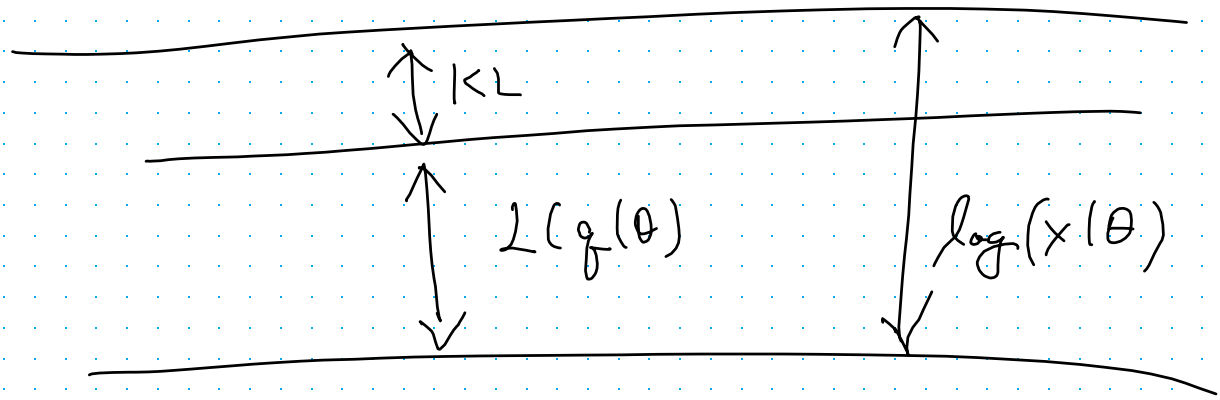
多峰性の分布の場合 点推定ではピーク情報を無視

$$\log p(X|\theta) = \mathcal{L}(q,\theta) + KL(q\|p)$$

下界

expectation, maximization

期待値化 最大化 アルゴリズム



$$\mathcal{L}(q,\theta) = \sum_Z q(Z) \log \left\{ \frac{p(Y,Z|\theta)}{q(Z)} \right\}$$

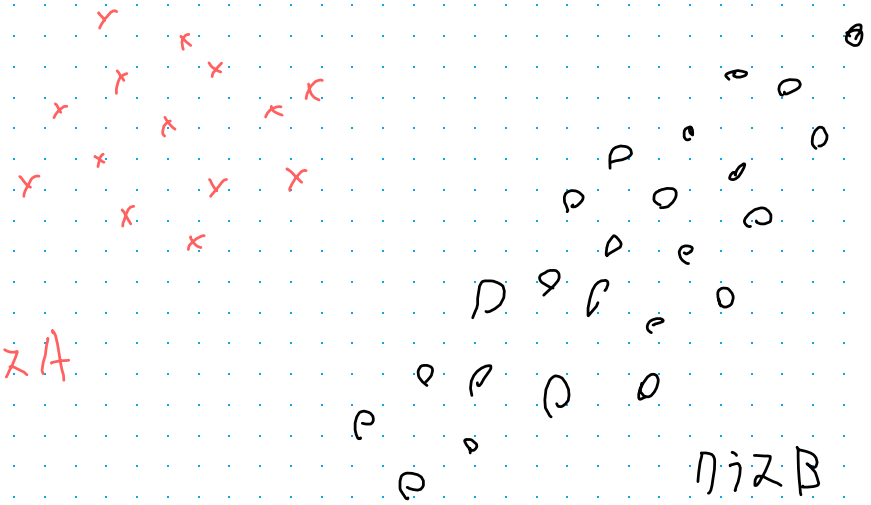
$$KL(q\|p) = \sum_Z q(Z) \log \left\{ \frac{q(Z)}{p(Z|X,\theta)} \right\}$$

E step θを固定し、 KL最小化する q を求める

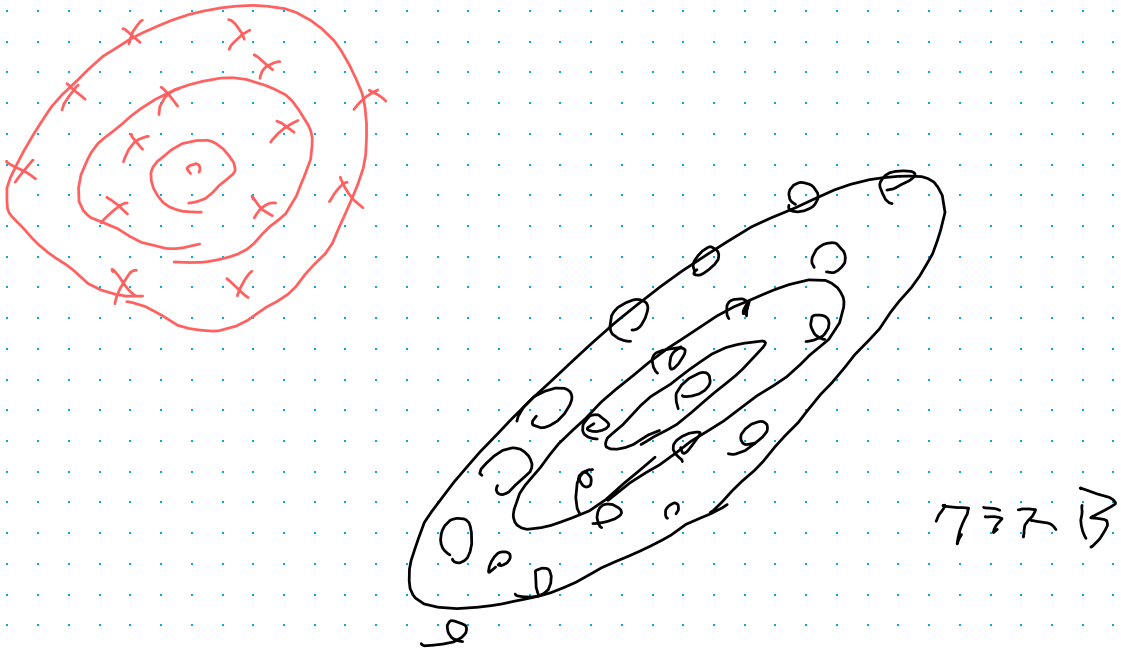
M step 下界をパラメータに関して最大化

Subject Date 4

Attendants



上記のようなデータが観測されたとすると



$P(y|x)$ を直接求めるのではなく、$P(x|y)P(y)$ を求めると

クラスを分類できるだけでなく、あるクラスに属する

擬似的なデータを作ることができるようになる

Subject 変分推論 Date 4-5

Attendants

$$\log p(X) = \mathcal{L}(q) + KL(q \| p)$$

対数モデルエビデンス

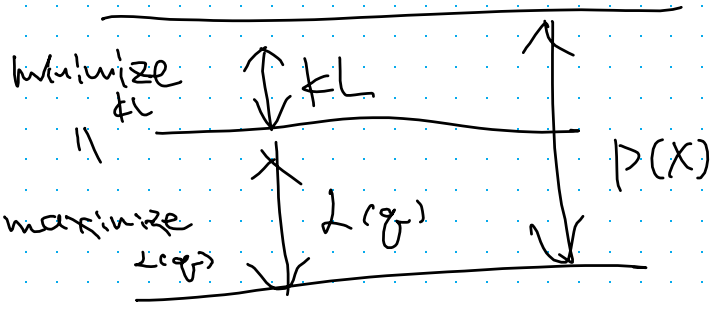
θを更新しても不変

$p(x)$はqに依存しないので

qに関してKL最小化すると

qに関して下界$\mathcal{L}(q)$を

最大化することにあたる



$p(x|z)$を求めたい

↓ 計算むずかしいので

$q(z)$で近似(下界)

↓

$KL(q \| p)$を最小化するqを

求める