情報認識
「最尤推定法」

- 講師:　杉山　将（計算工学専攻）
- 居室:　W8E-505
- 電子メール:　sugi@cs.titech.ac.jp

識別機の構成

- 最大事後確率則:

$$x \in \omega_k : k = \arg\max_i p(\omega_i \mid x)$$

- 事後確率 $p(\omega_i \mid x)$ は実際には未知.
- ベイズの定理を用いれば,

$$p(\omega_i \mid x) \propto \underbrace{p(x \mid \omega_i)}_{\text{条件付き確率}} \underbrace{p(\omega_i)}_{\text{事前確率}}$$

- 条件付き確率と事前確率も未知.

訓練標本

- 訓練標本が与えられる場合を考え，それらから条件付き確率と事前確率を推定することにする.
- 訓練標本(training sample) $\{x_i\}_{i=1}^{n}$:属するカテゴリが既知のパターン
- 訓練標本は独立に次のように生成されたと仮定.
  - カテゴリを事前確率 $p(\omega_i)$ に従ってランダムに選ぶ.
  - 選んだカテゴリに対して，パターンを条件付き確率 $p(x \mid \omega_i)$ に従ってランダムに取り出す.

事前確率の推定

- n_i :カテゴリ ω_i に属する訓練標本の数
- 事前確率は離散的な確率分布なので，単純にそのカテゴリに含まれる標本の割合で推定する.

$$\hat{p}(\omega_i) = \frac{n_i}{n}$$

条件付き確率の推定

- 条件付き確率は連続的な確率分布なので，事前確率のときのように単純には推定できない．
- パラメトリック法：
  - 最尤推定法
  - ベイズ推定法
  - 最大事後確率推定法
- ノンパラメトリック法：
  - 核密度推定法
  - 最近傍密度推定法

条件付き確率の推定（続き）

- 以後，簡単のため，条件付きでない確率密度関数 $p(x)$ を全訓練標本 $\{x_i\}_{i=1}^{n}$ から推定する問題を考える.
- カテゴリ ω_i に関する条件付き確率 $p(x|\omega_i)$ を推定するときは，ω_i に属する n_i 個の標本のみを用いればよい.

パラメトリック法

- θ :パラメータ
- Θ :パラメータの定義域
- パラメトリックモデル(parametric model)
  $q(x;\theta)$:有限次元のパラメータで記述された確率密度関数の族
- パラメトリック法(parametric method):パラメトリックモデルを用いて確率密度関数を推定する方法
- 例:ガウス関数(正規分布)の族を考え，平均と分散(共分散)のパラメータを推定する.

最尤推定法

- 最尤推定法(maximum likelihood estimation method): 手元にある訓練標本が最も生起しやすいようにパラメータ値を決める方法
- “最も尤もらしいようにパラメータの値を決める”
- 訓練標本 $\{x_i\}_{i=1}^{n}$ が生起する確率:

$$p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} q(x_i; \theta)$$

- 尤度(likelihood): これを θ の関数とみたもの

$$L(\theta) = \prod_{i=1}^{n} q(x_i; \theta)$$

最尤推定法(続き)

- 尤度を最大にするようにパラメータの値を決定.

$$\hat{\theta}_{ML} = \arg\max_{\theta} L(\theta)$$

$$L(\theta) = \prod_{i=1}^{n} q(x_i;\theta)$$

- $\hat{\theta}_{ML}$ を最尤推定量(maximum likelihood estimator)とよぶ.

最尤推定量のよさ

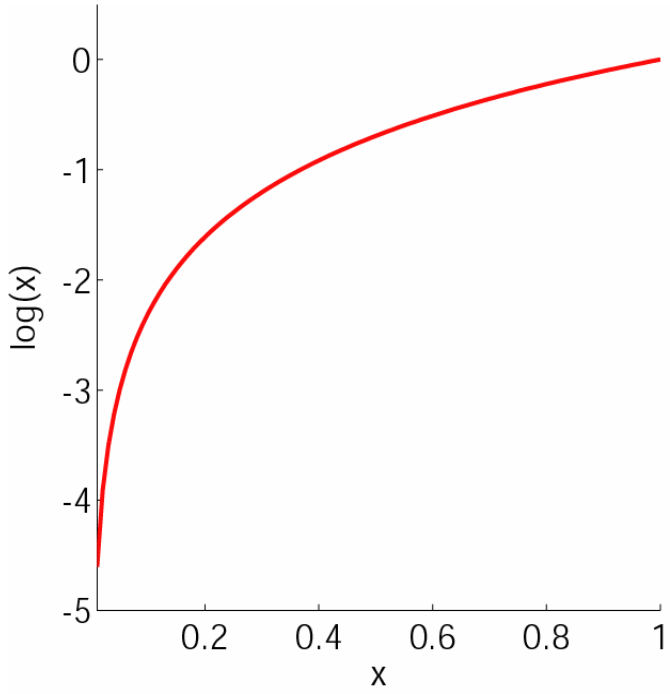
- θ_{true} :真の確率密度関数を表すパラメータ
- 一致性(consistency):

  $n \to \infty$ のとき, $\hat{\theta}_{ML} \to \theta_{true}$
- 漸近有効性(asymptotic efficiency):

  $n \to \infty$ のとき, $\hat{\theta}_{ML}$ の分散は漸近不偏推定量の中で最小(正確には, クラメル・ラオの下限を漸近的に達成する)

対数関数

- 対数(log)は，単調増加の関数.
- 対数をとってもその大小関係は変わらない.



対数尤度

- 対数をとれば積が和になることから，実際に最尤推定量を計算するときは，対数をとった尤度（対数尤度，log-likelihoodとよぶ）を用いた方が計算しやすいことが多い．

$$\hat{\theta}_{ML} = \arg\max_{\theta} \log L(\theta)$$

$$\log L(\theta) = \sum_{i=1}^{n} \log q(x_i;\theta)$$

- 最尤推定量は次の尤度方程式（likelihood equation)を満たす．

$$\left.\frac{\partial}{\partial\theta}\log L(\theta)\right|_{\theta=\hat{\theta}_{MLE}} = 0$$

ガウスモデル（正規分布）

- d 次元の確率ベクトル: $x = (x^{(1)}, x^{(2)}, \ldots, x^{(d)})^T$
- 2つのパラメータ:
  - d 次元ベクトル μ
  - d 次元正値行列 Σ

$$q(x;\mu,\Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

- ガウス分布の期待値，分散共分散行列

$$E[x] = \mu \qquad V[x] = \Sigma$$

ガウスモデルにおける最尤推定

ガウスモデルの最尤推定量

$$\hat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \hat{\Sigma}_{ML} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu}_{ML})(x_i - \hat{\mu}_{ML})^T$$

証明のヒント

- ベクトルでの微分の公式

$$\frac{\partial}{\partial \mu}\mu^T \Sigma \mu = 2\Sigma\mu \qquad \frac{\partial}{\partial \mu}\mu^T \Sigma x = \Sigma x$$

- 行列での微分の公式

$$\frac{\partial}{\partial \Sigma} x^T \Sigma^{-1} x = -\Sigma^{-1} x x^T \Sigma^{-1} \qquad \frac{\partial}{\partial \Sigma}\log|\Sigma| = \Sigma^{-1}$$

証明

■ 対数尤度

$$\log L(\theta) = \sum_{i=1}^{n}\left(-\frac{d}{2}\log 2\pi - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)\right)$$

$$= -\frac{nd}{2}\log 2\pi - \frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)$$

■ 尤度方程式

$$\frac{\partial}{\partial\mu}\log L(\theta) = -n\Sigma^{-1}\mu + \Sigma^{-1}\sum_{i=1}^{n}x_i = 0$$

$$\frac{\partial}{\partial\Sigma}\log L(\theta) = -\frac{n}{2}\Sigma^{-1} + \frac{1}{2}\sum_{i=1}^{n}\Sigma^{-1}(x_i-\mu)(x_i-\mu)^T\Sigma^{-1} = 0$$

小レポート

- 共分散がない，即ち，分散共分散行列が $\Sigma = diag(\sigma_i^2)$ で与えられるとき，σ_i^2 の最尤推定量を求めよ.
- 更に分散が等しいとき，即ち，分散共分散行列が $\Sigma = \sigma^2 I$ で与えられるとき，σ^2 の最尤推定量を求めよ.