情報認識
「最大事後確率推定法」

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

ベイズ推定法の近似

$$\hat{p}(x) = \int_{\Theta} q(x;\theta) \frac{\prod_{i=1}^{n} q(x_i;\theta)p(\theta)}{\int_{\Theta} \prod_{i=1}^{n} q(x_i;\theta')p(\theta')d\theta'} d\theta$$

- ラプラス近似を用いれば，比較的簡単にベイズ推定法が計算できる．
- しかし，ヘシアン行列を計算するのが面倒．
- もっと計算を簡単にできないか？

最大事後確率推定法

- **最大事後確率推定法** (maximum a posteriori estimation method, MAP**推定法**)：**パラメータに関して積分するのをやめて，事後確率を最大にするパラメータ1点を用いる**

$$\hat{p}(x) = q(x; \hat{\theta}_{MAP})$$

$$\hat{\theta}_{MAP} = \arg\max_{\theta} p(\theta \mid x_1, x_2, \ldots, x_n)$$

最大事後確率推定法の計算

- 最尤推定のときと同様に，対数をとったほうが計算しやすいことが多い．
- ベイズの定理より，

$$\log p(\theta \mid x_1, x_2, \ldots, x_n)$$

$$= \log p(x_1, x_2, \ldots, x_n \mid \theta) + \log p(\theta) + C$$

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

$$C = -\log p(x_1, x_2, \ldots, x_n)$$

最大事後確率推定法と最尤推定法

- MAP推定法はベイズ推定の近似解法.
- しかし，代表パラメータ $\hat{\theta}_{MAP}$ で分布を推定するため，ベイズ推定よりはむしろ最尤推定に近い.
- MAP推定法は，対数尤度に対数事前確率を加えたものを最大にする.
- そのためMAP法は，**罰則付き最尤推定法** (penalized MLE)とも呼ばれる.

$$\hat{\theta}_{MAP} = \arg\max_{\theta} \left(\underbrace{\sum_{i=1}^{n} \log q(x_i;\theta)}_{\text{対数尤度}} + \underbrace{\log p(\theta)}_{\text{対数事前確率}} \right)$$

事前確率の役割

- 最尤推定法は訓練標本が少ないとき精度がよくないことがある.
- MAP推定法では事前確率の大きい方に解が引っ張られるので, 訓練標本が少ない場合でも精度がよいことがある.
- 事前確率がわからない場合はどうすればよいか？
- データから事前確率を定める(経験ベイズ法)

演習

- 以下の設定でMAP推定量と最尤推定量を求め，その違いを論ぜよ．
  - モデル：共分散がなく分散が等しいガウスモデル

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

  - 標本： $\{x_i\}_{i=1}^n$
  - 事前確率：指数分布

$$p(\mu,\sigma^2)=\beta\exp(-\beta\sigma^2)\qquad \beta>0$$

（小さい σ^2 の方が確率が大きい）

小レポート

- 「情報認識」の授業の約半分が終了した.
- 課題：ここまでの**授業を評価**せよ.
  - 良いところは具体的にどういうところが良かったか記述し，悪いところは**どうすれば改善するか**も考えよ.
  - 授業で扱っているトピック，授業の進め方，講義内容の難易度，ウェブで公開している資料，宿題などについての意見を述べよ.
  - その他，授業内容の改善に役立つ意見・感想を述べよ.
- このレポートは他のレポートと同様，成績評価に加味するので，真剣に書くこと.