言語処理のための機械学習入門

1 .5 パラメータ推定法

河野　和平

パラメータ推定法

- 言語処理研究
  - 言語データの性質から確率分布の種類を仮定
  - パラメータの値をデータから決定

- パラメータ推定問題
  - 与えられたデータがある種類の確率分布から生成されたと仮定する。
  - パラメータの値は分からない。

尤度

- 独立に同一の確率分布に従うデータ(i.i.d.)
  - 確率変数 X のサンプルデータ $D = \left\{x^{(1)}, \cdots, x^{(N)}\right\}$
  - 生成確率(尤度)

$$P(D) = \prod_{x^{(i)} \in D} p\left(x^{(i)}\right)$$

$$\begin{aligned}\log P(D) &= \log \prod_{x^{(i)} \in D} p\left(x^{(i)}\right) \\ &= \sum_{x^{(i)} \in D} \log p\left(x^{(i)}\right) \\ &= \sum_{x} n_x \log p(x)\end{aligned}$$

最尤推定

i.i.d.であるデータ $D=\left\{x^{(1)},\cdots,x^{(N)}\right\}$ が与えられたとき、対数尤度が最も高くなるようにパラメータを決定する。

ポアソン分布（パラメータ μ ）

$$\log P(D)=\sum_{x^{(i)}\in D}\log\frac{\mu^{x^{(i)}}}{x^{(i)}!}e^{-\mu}$$

最尤推定

ポアソン分布（パラメータμ）

$$\log P(D) = \sum_{x^{(i)} \in D} \log \frac{\mu^{x^{(i)}}}{x^{(i)}!} e^{-\mu}$$

$$= \sum_{x^{(i)} \in D} \log \frac{\mu^{x^{(i)}}}{x^{(i)}!} + \log e^{-\mu}$$

$$\frac{\partial \log P(D)}{\partial \mu} = \sum_{x^{(i)} \in D} \left(x^{(i)} \frac{1}{\mu} - 1 \right)$$

$$\sum_{x^{(i)} \in D} x^{(i)} \frac{1}{\mu} = \sum_{x^{(i)} \in D} 1 = |\mathrm{D}|$$

$$\mu = \frac{\sum_{x^{(i)} \in D} x^{(i)}}{|\mathrm{D}|}$$

最大事後確率推定

パラメータがどんな値をとりやすいか事前に分かっていると仮定する。

パラメータθの確率分布 $P(\theta)$ ･･･<u>事前確率分布</u>

データ D が与えられたときの
パラメータθの確率分布 $P(\theta|D)$ ･･･<u>事後確率分布</u>

事後確率 $P(\theta|D)$ が最大になるように
パラメータを決定

最大事後確率推定

- 事後確率の最大化

$$\arg\max_{\theta} P(\theta|D) = \arg\max_{\theta} \frac{P(\theta)\cdot P(D|\theta)}{P(D)}$$

$$= \arg\max_{\theta} P(\theta)\cdot P(D|\theta)$$

を最大化するようなパラメータ θ を求める。

簡単のため、対数をとって計算する。

$$\begin{aligned}
&\log P(\theta)\cdot P(D|\theta) \\
&= \log P(\theta) + \log P(D|\theta) \\
&= \log P(\theta) + \sum_{x^{(i)}\in D} \log P\left(x^{(i)}|\theta\right)
\end{aligned}$$