平成24年度 東京大学理学部情報科学科

月曜2限10:30〜12:00

理7-214

知識処理論

2012年4月23日(月)

井元清哉

東京大学医科学研究所

ヒトゲノム解析センター

DNA情報解析分野

http://bonsai.hgc.jp/~imoto

imoto@ims.u-tokyo.ac.jp

正則化法1

β の推定については、対数尤度

$$l(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X\beta})^T(\boldsymbol{y} - \boldsymbol{X\beta})$$

の最大化は、残差2乗和

$$S(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X\beta})^T(\boldsymbol{y} - \boldsymbol{X\beta})$$

の最小化と同値

$$\hat{\boldsymbol{\beta}} = \min{}_{\beta} S(\boldsymbol{\beta}) \Rightarrow \hat{\boldsymbol{\beta}} = \underline{(\boldsymbol{X}^T\boldsymbol{X})}^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

正則化法2

残差2乗和を修正する

$$S_\lambda(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + \lambda\boldsymbol{\beta}^T\boldsymbol{\beta}$$

$\hat{\boldsymbol{\beta}}_{\mathrm{P}} = \min_{\boldsymbol{\beta}} S_\lambda(\boldsymbol{\beta})$ を求めてみましょう

正則化法3

$\hat{\boldsymbol{\beta}}_{\mathrm{P}} = \min_{\boldsymbol{\beta}} S_\lambda(\boldsymbol{\beta})$ は次式で与えられる:

$$\hat{\boldsymbol{\beta}}_{\mathrm{P}} = (\underline{\boldsymbol{X}^T\boldsymbol{X} + \lambda I})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

$\boldsymbol{X}^T\boldsymbol{X}$ の対角成分に λ を足している ⇒ 正則化

λ の値を適切に決める必要がある。それは後日の講義で。

$S_\lambda(\boldsymbol{\beta})$ は罰則付き残差2乗和(Penalized sum of squared residuals)と呼ばれる。
$\hat{\boldsymbol{\beta}}_{\mathrm{P}}$ は Ridge estimator と呼ばれる。

正則化法と最尤法

$$S_\lambda(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X\beta})^T(\boldsymbol{y} - \boldsymbol{X\beta}) + \lambda\boldsymbol{\beta}^T\boldsymbol{\beta}$$

$$\Rightarrow -\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X\beta})^T(\boldsymbol{y} - \boldsymbol{X\beta}) - \frac{\lambda}{2}\boldsymbol{\beta}^T\boldsymbol{\beta}$$

$$\Rightarrow \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X\beta})^T(\boldsymbol{y} - \boldsymbol{X\beta})\right\}\cdot\left(\frac{\lambda}{2\pi}\right)^{(p+1)/2}\exp\left(-\frac{\lambda}{2}\boldsymbol{\beta}^T\boldsymbol{\beta}\right)$$

$$= \underline{L(\boldsymbol{\beta},\sigma^2)}\cdot\underline{p(\boldsymbol{\beta})}$$

尤度　　　この項の解釈は？

ベイズ統計学から解釈ができる

ベイズ統計学

データ: $x_1, x_2, \ldots, x_n \sim X$

統計モデル: $f(x|\boldsymbol{\theta})$

尤度: $L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta})$

ベイズ統計学における推測は $\boldsymbol{\theta}$ の事後分布

$$p(\boldsymbol{\theta}\,|\,\boldsymbol{x}) = \frac{L(\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{x})}$$

によって行われる。

ベイズの定理

事象 A, B に対して　$\Pr(A \mid B) = \dfrac{\Pr(B \mid A)\Pr(A)}{\Pr(B)}$　が成り立つ。

事象 $A_1, A_2, \ldots, A_m$ が $A_i \cap A_j = \phi, A_1 \cup A_2 \cup \cdots \cup A_m = \Omega$ を満たすとき

$$\Pr(B) = \Pr(B \mid A_1)\Pr(A_1) + \Pr(B \mid A_2)\Pr(A_2) + \cdots + \Pr(B \mid A_m)\Pr(A_m)$$

が成り立つ。

これらを合わせると

$$\Pr(A_i \mid B) = \frac{\Pr(B \mid A_i)\Pr(A_i)}{\sum_{j=1}^{m} \Pr(B \mid A_j)\Pr(A_j)}$$

が成り立つ。これはベイズの定理と呼ばれる。

結果から原因を探る

確率変数: X, Y

X は結果 Y の原因である:

$\Pr(Y \mid X)$

は原因が与えられた下での結果が起こる確率を表す。

(例) X:夜更かし、Y:朝寝坊

(例) X:アスベスト、Y:中皮腫

(例) X:SNP、Y:癌　　AGGTCCA $\genfrac{}{}{0pt}{}{\text{A}}{\text{G}}$ TGAAAA

要するに、$\Pr(Y \mid X)$ は普通に想像されうる確率。

今、X と Y を入れ替えた確率 $\Pr(X \mid Y)$ を考える。

この確率は、結果が与えられた下での原因の起こる確率である。

複数の原因 $X_1, X_2, \ldots, X_m$ が想定される状況を考える:このとき、

$\Pr(X_1 \mid Y),\ \Pr(X_2 \mid Y), \ldots, \Pr(X_m \mid Y)$

を計算し、その最大値 $\max_X \Pr(X \mid Y)$ を与える X を探す。

結果 Y が起こったという観測データに基づく
原因 X を特定するための科学的方法論を与えている.

θの事後分布からの推測

$$p(\boldsymbol{\theta} \mid \boldsymbol{x}) = \frac{L(\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{x})} = \frac{\text{尤度} \times \theta\text{の事前分布}}{\text{規格化定数}}$$

パラメータ θの推定

$$\hat{\boldsymbol{\theta}}_{\mathrm{MAP}} = \mathrm{argmax}_{\boldsymbol{\theta}}\ p(\boldsymbol{\theta} \mid \boldsymbol{x})$$ 事後確率最大化（Maximum *a posteriori*）

$$\hat{\boldsymbol{\theta}}_{\mathrm{Bayes}} = E_{p(\boldsymbol{\theta} \mid \boldsymbol{x})}[\boldsymbol{\theta}]$$ 期待値

新しいデータ z の分布 $p(z \mid \boldsymbol{x})$（予測分布）

MAP推定では $p(z \mid \boldsymbol{x}) = p(z \mid \hat{\boldsymbol{\theta}}_{\mathrm{MAP}})$

ベイズ推定では $p(z \mid \boldsymbol{x}) = \int p(z \mid \boldsymbol{\theta})p(\boldsymbol{\theta} \mid \boldsymbol{x})d\boldsymbol{\theta}$

知識のアップデート

パラメータ θに
関する事前の知識

パラメータ θの
事前分布

$p(\boldsymbol{\theta})$

データ $\boldsymbol{x}$ の情報を
加味したパラメータ θの分布

パラメータ θの
事前分布

$p(\boldsymbol{\theta} \mid \boldsymbol{x})$

得られたデータ $\boldsymbol{x}$ の
情報を θの空間で表現

$L(\boldsymbol{\theta})$

$$\boxed{p(\boldsymbol{\theta} \mid \boldsymbol{x}) = \frac{L(\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{x})}}$$

事前分布、尤度、事後分布：例