パターン認識と機械学習
1.2.5 曲線フィッティング再訪

吉田拓夢

曲線フィッティング問題

・N個の入力値の訓練データ集合 $\mathbf{x} = (x_1, \ldots, x_N)^{\mathrm{T}}$
　それに対応する目標値 $\mathbf{t} = (t_1, \ldots, t_N)^{\mathrm{T}}$
→新たな入力値xに対する目標変数tの予測

・　目標変数の値に関する不確実性→確率分布
　xに対するtは平均が多項式曲線y(x,**w**)に等しい
　ガウス分布に従う

$$p(t|x, \mathbf{w}, \beta) = \mathcal{N}(t|y(x, \mathbf{w}), \beta^{-1})$$

β: 分布の逆分散に相当するパラメータ

最尤推定

- 訓練データ$\{\mathbf{x}, \mathbf{t}\}$を用いて未知パラメータ**w**,βを求める
  データが分布から独立に取られるとして尤度関数は

$$p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}(t_n|y(x_n, \mathbf{w}), \beta^{-1})$$

- 尤度関数の最大化が便利
  ガウス分布の形を置きかえると対数尤度関数は

$$\ln p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = -\frac{\beta}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{N}{2}\ln\beta - \frac{N}{2}\ln(2\pi)$$

最尤推定

- 最尤解で定まる多項式の係数 $\mathbf{w}_{\mathrm{ML}}$ (**w**の最大化)

$$-\frac{\beta}{2}\sum_{n=1}^{N}\{y(x_n,\mathbf{w})-t_n\}^2+\frac{N}{2}\ln\beta-\frac{N}{2}\ln(2\pi)$$

$$-\frac{\beta}{2}\sum_{n=1}^{N}\{y(x_n,\mathbf{w})-t_n\}^2$$ ←**w**に依存しない項を無視

$$-\frac{1}{2}\sum_{n=1}^{N}\{y(x_n,\mathbf{w})-t_n\}^2$$ ←正の定数倍で**w**の最大値の位置は変わらない

$$\frac{1}{2}\sum_{n=1}^{N}\{y(x_n,\mathbf{w})-t_n\}^2$$ ←-1倍：(最大化→最小化)

- **w**を決める観点で**二乗和誤差**の最小化と等価
- **二乗和誤差関数**←尤度の最大化の結果（ノイズがガウス分布に従う仮定の下で）

最尤推定

- 条件付きガウス分布の精度パラメータβ

$$\frac{1}{\beta_{\mathrm{ML}}} = \frac{1}{N}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}_{\mathrm{ML}}) - t_n\}^2$$

予測分布

- パラメータw,βの決定→xの新たな値の予測
- 確率モデルで定式化→予測分布という形でtの確率分布を与えられる

$$p(t|x, \mathbf{w}_{\mathrm{ML}}, \beta_{\mathrm{ML}}) = \mathcal{N}(t|y(x, \mathbf{w}_{\mathrm{ML}}), \beta_{\mathrm{ML}}^{-1})$$

事前分布の導入

- 簡便のため、以下のガウス分布を考える

$$p(\mathbf{w}|\alpha) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha^{-1}\mathbf{I}) = \left(\frac{\alpha}{2\pi}\right)^{(M+1)/2} exp\left\{-\frac{\alpha}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}\right\}$$

α：分布の精度パラメータ←**超パラメータ(hyperparameter)**

M+1：M次多項式に対する**w**の要素数

- ベイズの定理からwの事後分布は事前分布と尤度関数との積に比例

$$p(\mathbf{w}|\mathbf{x}, \mathbf{t}, \alpha, \beta) \propto p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta)p(\mathbf{w}|\alpha)$$

- 与えられたデータに基づく最も確からしい**w**の値を見つける。

(事後分布を最大化する**w**を決める)

→**最大事後確率推定**(**ma**ximum **p**osterior)/MAP推定

最大事後確率推定

- 事後確率の最大値は以下の最小値として与えられる

$$\frac{\beta}{2}\sum_{n=1}^{N}\{y(x_n,\mathbf{w})-t_n\}^2+\frac{\alpha}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}$$

{ $p(\mathbf{t}|\mathbf{x},\mathbf{w},\beta)p(\mathbf{w}|\alpha)$を対数をとって符号を反転 }

→正則化された二乗和誤差の最小化と等価

$$\widetilde{E}(\mathbf{w})=\frac{1}{2}\sum_{n=1}^{N}\{y(x_n,\mathbf{w})-t_n\}^2+\frac{\lambda}{2}\parallel\mathbf{w}\parallel^2$$

正規化パラメータは$\lambda=\alpha/\beta$