パターン認識と機械学習
1.5.5 回帰のための損失関数

茂木 哲矢

2008 年 5 月 16 日

目次

クラス分類問題

例．癌の診断

$$
\begin{array}{cc}
 & \begin{array}{cc}\text{癌} & \text{正常}\end{array} \\
\begin{array}{c}\text{癌} \\ \text{正常}\end{array} & \begin{pmatrix} 0 & 1000 \\ 1 & 0 \end{pmatrix}
\end{array}
$$

- 癌の人を正常と診断する．$\rightarrow$ まずいので，損失大．
- 正常の人を癌と診断する．$\rightarrow$ 再検査で癌でないと分かるため，損失小．

曲線フィッティングって？

訓練データ x, t から，関数 $y(x)$ を作り，新しい x を入力したときの $y(x)$ を予測すること．

$$y(x, \boldsymbol{w}) = \sum_{j=0}^{M} w_j x^j$$

これまでも，

- 二乗和誤差関数
- 正則化された二乗和誤差関数

のように $y(x, w)$ の求め方を見てきた．

回帰問題の損失関数

決定段階:
x **に対する** t **に適当な** $y(x)$ **を選ぶこと**

- **平均損失（期待損失）**

$$\mathbb{E}[L] = \iint L(t, y(\boldsymbol{x}))p(\boldsymbol{x}, t)d\boldsymbol{x}dt$$

- **損失関数=二乗誤差 の場合**

$$\mathbb{E}[L] = \iint \{y(\boldsymbol{x}) - t\}^2 p(\boldsymbol{x}, t)d\boldsymbol{x}dt$$

$y(\boldsymbol{x})$ を求める

$\mathbb{E}[L]$ を最小化する $y(\boldsymbol{x})$ を求める．$\rightarrow$ 変分法を使う．

$$\frac{\delta \mathbb{E}[L]}{\delta y(\boldsymbol{x})} = 2 \int \{y(\boldsymbol{x}) - t\} p(\boldsymbol{x}, t) dt = 0$$

確率の加法・乗法定理[1]より，

$$y(\boldsymbol{x}) = \frac{\int t p(\boldsymbol{x}, t) dt}{p(\boldsymbol{x})} = \int t p(t|\boldsymbol{x}) dt = \mathbb{E}_t[t|\boldsymbol{x}]$$

条件付き期待値 $\mathbb{E}_t[t|\boldsymbol{x}]$:
　回帰関数という．

[1]加法定理 $\int p(x, y) dy = p(x)$，乗法定理 $p(x, y) = p(y|x)p(x)$．

変分法って

- 汎関数（関数 $y(x)$ を入力としてもつ関数）の微分のようなもの．
- 汎関数

$$F[y] = \int_{x_1}^{x_2} G(y(x), y'(x), x) dx.$$

の解は，オイラー − ラグランジュ方程式を解くことで得られる．

$$\frac{\delta F}{\delta y(x)} = \frac{\partial G}{\partial y} - \frac{d}{dx}\left(\frac{\partial G}{\partial y'}\right) = 0.$$

別のアプローチで $y(x)$ を求める

$$\begin{aligned}\{y(\boldsymbol{x})-t\}^2 &= \{y(\boldsymbol{x})-\mathbb{E}[t|\boldsymbol{x}]+\mathbb{E}[t|\boldsymbol{x}]-t\}^2\\ &= \{y(\boldsymbol{x})-\mathbb{E}[t|\boldsymbol{x}]\}^2\\ &\quad +2\{y(\boldsymbol{x})-\mathbb{E}[t|\boldsymbol{x}]\}\{\mathbb{E}[t|\boldsymbol{x}]-t\}\\ &\quad +\{\mathbb{E}[t|\boldsymbol{x}]-t\}^2\end{aligned}$$

平均損失は，

$$\mathbb{E}[L]=\int\{y(\boldsymbol{x})-\mathbb{E}[t|\boldsymbol{x}]\}^2p(\boldsymbol{x})d\boldsymbol{x}+\int\left\{\int\{t-\mathbb{E}[t|\boldsymbol{x}]\}^2p(t|\boldsymbol{x})dt\right\}p(\boldsymbol{x})d\boldsymbol{x}$$

- **第 1 項目** 0 **で** $\mathbb{E}[L]$ **が最小．** $\rightarrow y(\boldsymbol{x})=\mathbb{E}[t|\boldsymbol{x}]$**．**
- **第 2 項目** **データが元から持つ変動．** $\rightarrow$ **ノイズ．**

t による積分

第 1 項目 $\int\{y(\boldsymbol{x})-\mathbb{E}[t|\boldsymbol{x}]\}^2\int p(\boldsymbol{x},t)dtd\boldsymbol{x}$

$$= \int\{y(\boldsymbol{x})-\mathbb{E}[t|\boldsymbol{x}]\}^2 p(\boldsymbol{x})d\boldsymbol{x}$$

第 2 項目 $\int 2\{y(\boldsymbol{x})-\mathbb{E}[t|\boldsymbol{x}]\}\int\{\mathbb{E}[t|\boldsymbol{x}]-t\}p(\boldsymbol{x},t)dtd\boldsymbol{x}$ **より，**

$$\int\{\mathbb{E}[t|\boldsymbol{x}]-t\}p(\boldsymbol{x},t)dt$$
$$= \mathbb{E}[t|\boldsymbol{x}]\int p(\boldsymbol{x},t)dt-\int tp(t|\boldsymbol{x})p(\boldsymbol{x})dt$$
$$= \mathbb{E}[t|\boldsymbol{x}]p(\boldsymbol{x})-\mathbb{E}[t|\boldsymbol{x}]p(\boldsymbol{x})=0$$

第 3 項目 $\iint\{\mathbb{E}[t|\boldsymbol{x}]-t\}^2 p(\boldsymbol{x},t)dtd\boldsymbol{x}$

$$= \iint\{\mathbb{E}[t|\boldsymbol{x}]-t\}^2 p(t|\boldsymbol{x})p(\boldsymbol{x})dtd\boldsymbol{x}$$

問題解決の 3 つのアプローチ

1.
    - 同時分布 $p(x, t)$ を推定する問題を解く．
    - 条件付き密度 $p(t|x)$ を求める．
    - 条件付き平均を求める．
2.
    - 条件付き密度 $p(t|x)$ を求める．
    - 条件付き平均を求める．
3. 回帰関数を直接計算する．

損失関数の補足

逆問題[2]を扱う場合に結果が悪くなることがある．
回帰の損失関数は二乗損失のみでない．

二乗誤差の一般化．$\rightarrow$ ミンコフスキー損失．

$$\mathbb{E}[L_q] = \iint |y(\boldsymbol{x}) - t|^q p(\boldsymbol{x}, t) d\boldsymbol{x} dt$$

$\mathbb{E}[L_q]$ の最小値は

$q = 2$ 条件付き期待値

$q = 1$ 条件付きメディアン

$q \rightarrow 0$ 条件付きモード

[2]出力から入力を求めるような問題