演習問題 1.27

回帰の問題で、損失関数 L_q が式 (1.91) で与えられるときの期待損失を考える。 $y(\mathbf{x})$ が $\mathrm{E}[L_q]$ を最小化するために満たすべき条件を書き下せ。$q=1$ に対しては解が条件付きメディアンになる。つまり、$t < y(\mathbf{x})$ となる確率質量と $t \geq y(\mathbf{x})$ となる確率質量は等しいことを示せ。また、$q \to 0$ に対する L_q の期待損失を最小にするのは、条件付きモードになること、つまり、関数 $y(\mathbf{x})$ が、各 $\mathbf{x}$ に対して、$p(t \mid \mathbf{x})$ を最大にする t の値に等しくなることを示せ。

[ミンコフスキー損失]

二乗誤差を単純に一般化したものである。

$$\mathrm{E}[L_q] = \int\int |y(\mathbf{x}) - t|^q p(\mathbf{x}, t)\, d\mathbf{x}\, dt$$

… (1.91)

[解]

$q=1$ と $q \to 0$ のときに対して、$y(\mathbf{x})$ が $\mathrm{E}[L_q]$ を最小化するために満たすべき条件を書き下す。そのためには、$\mathrm{E}[L_q]$ を最小化するために、式 (1.91) を $y(\mathbf{x})$ について変分する必要があるが、この前に式 (1.91) を $\mathbf{x}$ について整理しておく。すると、式 (1.91) は、確率の乗法定理より、

$$\mathrm{E}[L_q] = \int\int |y(\mathbf{x}) - t|^q p(\mathbf{x}, t)\, d\mathbf{x}\, dt$$

$$= \int\int |y(\mathbf{x}) - t|^q p(t \mid \mathbf{x})\, p(\mathbf{x})\, d\mathbf{x}\, dt$$

$$= \int\int |y(\mathbf{x}) - t|^q p(\mathbf{x})\, d\mathbf{x}\, p(t \mid \mathbf{x})\, dt$$

と展開できるので、この式を $y(\mathbf{x})$ について変分すると、

$$\frac{\partial\, \mathrm{E}[L_q]}{\partial\, y(\mathbf{x})} = \frac{\partial \left\{ \int\int |y(\mathbf{x}) - t|^q p(\mathbf{x})\, d\mathbf{x}\, p(t \mid \mathbf{x})\, dt \right\}}{\partial\, y(\mathbf{x})}$$

$$= \int q\, |y(\mathbf{x}) - t|^{q-1} p(t \mid \mathbf{x})\, dt = 0$$

… ※

となる。

まず、$q=1$ のとき、解が条件付きメディアンとなること、つまり、$t < y(\mathbf{x})$ となる確率質量と $t \geq y(\mathbf{x})$ となる確率質量が等しくなることを示す。このとき、式 ※ は、

$$\int_{-\infty}^{y(\mathbf{x})} q\, |y(\mathbf{x}) - t|^{q-1} p(t \mid \mathbf{x})\, dt - \int_{y(\mathbf{x})}^{\infty} q\, |y(\mathbf{x}) - t|^{q-1} p(t \mid \mathbf{x})\, dt = 0$$

の積分区間で分けることができる。ただし、$t \geq y(\mathbf{x})$ となる確率質量のとき、$y(\mathbf{x}) - t \leq 0$、かつ $q = 1$ のとき、$| \, y(\mathbf{x}) - t \, |^{q} \leq 0$ より、上記の式の第二項の符号がマイナスになっていることに注意する。すると、上記の式は、$q = 1$ より、

$$\int_{-\infty}^{y(\mathbf{x})} 1 \cdot | \, y(\mathbf{x}) - t \, |^{1-1} \, p(\, t \mid \mathbf{x} \,) \, dt = \int_{y(\mathbf{x})}^{\infty} 1 \cdot | \, y(\mathbf{x}) - t \, |^{1-1} \, p(\, t \mid \mathbf{x} \,) \, dt$$

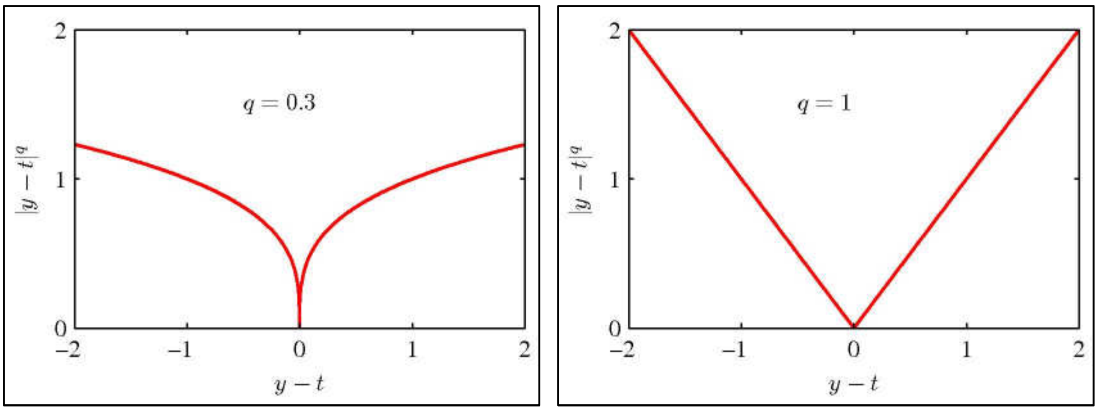
$$\int_{-\infty}^{y(\mathbf{x})} p(\, t \mid \mathbf{x} \,) \, dt = \int_{y(\mathbf{x})}^{\infty} p(\, t \mid \mathbf{x} \,) \, dt$$

となる。これより、$t < y(\mathbf{x})$ となる確率質量と $t \geq y(\mathbf{x})$ となる確率質量が等しくなるので、$q = 1$ のとき、解が t の条件付きメディアンとなることが示された。

また、$q \to 0$ に対する L_q の期待損失を最小にするのは、条件付きモードになること、つまり、関数 $y(\mathbf{x})$ が、各 $\mathbf{x}$ に対して、$p(\, t \mid \mathbf{x} \,)$ を最大にする t の値に等しくなることを示す。$q \to 0$ のとき、式 (1.91) より、$y(\mathbf{x}) = t$ のときを除いて、$\lim_{q \to 0} | \, y(\mathbf{x}) - t \, |^{q} = 1$ となる。確率密度 $p(\, t \,)$ は 1 に正規化されているので、式 (1.91) は 1 に収束する。また、$y(\mathbf{x}) = t$ のときには、$| \, y(\mathbf{x}) - t \, |^{q}$ は 0 となる。ただし、$y(\mathbf{x}) - t \to 0$ のとき、$| \, y(\mathbf{x}) - t \, |^{q}$ はV 字刻みの形で 0 に収束していく。つまり、式 (1.91) は、$q \to 0$ において、

$$\begin{cases} \lim\limits_{q \to 0} \mathrm{E}[\, L_q \,] = \displaystyle\iint | \, y(\mathbf{x}) - t \, |^{q} \, p(\mathbf{x}, t) \, d\mathbf{x} \, dt = 1 & (\, y(\mathbf{x}) \neq t \text{ のとき} \,) \\ \lim\limits_{q \to 0} \mathrm{E}[\, L_q \,] = \displaystyle\iint | \, y(\mathbf{x}) - t \, |^{q} \, p(\mathbf{x}, t) \, d\mathbf{x} \, dt = 0 & (\, y(\mathbf{x}) = t \text{ のとき} \,) \end{cases}$$

となることがわかる。このことから、確率密度 $p(\, t \,)$ の最大値に一致するように、V 字の谷の位置を選択することにより、期待損失 (1.91) において、値 0 を得ることができる。つまり、$q \to 0$ に対する L_q の期待損失を最小にするのは、条件付きモードであることがわかる。



図：$q = 1$，$q \to 0$ の値に対する $L_q = | \, y(\mathbf{x}) - t \, |^{q}$ のプロット
$\mathrm{E}[\, L_q \,]$ が最小となるのは、$q = 1$ では条件付きメディアン、$q \to 0$ では条件付きモードのときである。