Data Science for Actuaries (ACT6100)

Arthur Charpentier

Supervisé # 1 (Concepts Fondamentaux - 3)

automne 2020

https://github.com/freakonometrics/ACT6100/

ℓ_2 loss function

Let $\boldsymbol{y} \in \mathbb{R}^d$, $\overline{y} = \underset{m\in\mathbb{R}}{\text{argmin}} \left\{ \sum_{i=1}^{n} \frac{1}{n} \big[\underbrace{y_i - m}_{\varepsilon_i}\big]^2 \right\}$.

It is the empirical version of

$$\mathbb{E}[Y] = \underset{m\in\mathbb{R}}{\text{argmin}} \left\{ \int \big[\underbrace{y - m}_{\varepsilon}\big]^2 dF(y) \right\} = \underset{m\in\mathbb{R}}{\text{argmin}} \left\{ \mathbb{E}\big[\|\underbrace{Y - m}_{\varepsilon}\|_{\ell_2}\big] \right\}$$

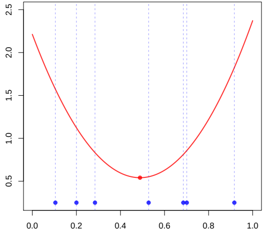
where Y is a random variable.

Thus, $\underset{m:\mathbb{R}^k\to\mathbb{R}}{\text{argmin}} \left\{ \sum_{i=1}^{n} \frac{1}{n} \big[\underbrace{y_i - m(\boldsymbol{x}_i)}_{\varepsilon_i}\big]^2 \right\}$ is the empirical version of $\mathbb{E}[Y|\boldsymbol{X} = \boldsymbol{x}]$.

See Legendre (1805) and Gauß (1809)

ℓ_2 loss function

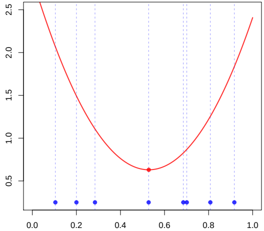
Sketch of proof: (1) Let $h(x)=\displaystyle\sum_{i=1}^d (x-y_i)^2$,



$$h'(x)=\sum_{i=1}^d 2(x-y_i)$$

and the FOC yields $x=\displaystyle\frac{1}{n}\sum_{i=1}^d y_i=\overline{y}$.

(2) If Y is continuous, let $h(x)=\displaystyle\int_{\mathbb{R}}(x-y)f(y)dy$ and $h'(x)$ is



$$\frac{\partial}{\partial x}\int_{\mathbb{R}}(x-y)^2 f(y)dy=\int_{\mathbb{R}}\frac{\partial}{\partial x}(x-y)^2 f(y)dy$$

i.e. $x=\displaystyle\int_{\mathbb{R}} xf(y)dy=\int_{\mathbb{R}} yf(y)dy=\mathbb{E}[Y]$

ℓ_1 loss function

Let $\boldsymbol{y} \in \mathbb{R}^d$, $\text{median}[\boldsymbol{y}] \in \underset{m \in \mathbb{R}}{\text{argmin}} \left\{ \sum_{i=1}^{n} \frac{1}{n} \big|\underbrace{y_i - m}_{\varepsilon_i}\big| \right\}$.

It is the empirical version of

$$\text{median}[Y] \in \underset{m \in \mathbb{R}}{\text{argmin}} \left\{ \int \big|\underbrace{y - m}_{\varepsilon}\big| dF(y) \right\} = \underset{m \in \mathbb{R}}{\text{argmin}} \left\{ \mathbb{E}\big[\|\underbrace{Y - m}_{\varepsilon}\|_{\ell_1}\big] \right\}$$

where Y is a random variable, $\mathbb{P}[Y \leq \text{median}[Y]] \geq \dfrac{1}{2}$ and $\mathbb{P}[Y \geq \text{median}[Y]] \geq \dfrac{1}{2}$.

$\underset{m:\mathbb{R}^k \to \mathbb{R}}{\text{argmin}} \left\{ \sum_{i=1}^{n} \frac{1}{n} \big|\underbrace{y_i - m(\boldsymbol{x}_i)}_{\varepsilon_i}\big| \right\}$ is the empirical version of $\text{median}[Y|\boldsymbol{X} = \boldsymbol{x}]$.

See Boscovich (1757) and Laplace (1793).

ℓ_1 loss function

Sketch of proof: (1) Let $h(x) = \displaystyle\sum_{i=1}^{d} |x - y_i|$

(2) If F is absolutely continuous, $dF(x) = f(x)dx$, and the median m is solution of $\displaystyle\int_{-\infty}^{m} f(x)dx = \frac{1}{2}$.

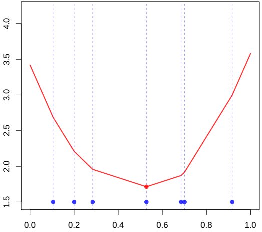
Set $h(y) = \displaystyle\int_{-\infty}^{+\infty} |x - y| f(x)dx$

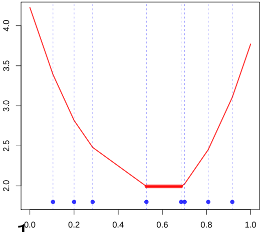
$$= \int_{-\infty}^{y} (-x + y) f(x)dx + \int_{y}^{+\infty} (x - y) f(x)dx$$

Then $\quad h'(y) = \displaystyle\int_{-\infty}^{y} f(x)dx - \int_{y}^{+\infty} f(x)dx$,

and FOC yields

$$\int_{-\infty}^{y} f(x)dx = \int_{y}^{+\infty} f(x)dx = 1 - \int_{-\infty}^{y} f(x)dx = \frac{1}{2}$$





Quantile (asymmetric) loss function

If $\tau \in (0,1)$, $Q(\tau) = F^{-1}(\tau) = \inf\{y : F(y) \geq \tau\}$. Let

$$q = \underset{u}{\text{argmin}} \left\{ (\tau - 1) \int_{-\infty}^{u} (y-u) dF(y) + \tau \int_{u}^{\infty} (y-u) dF(y) \right\}.$$

The first order condition is

$$0 = (1-\tau) \int_{-\infty}^{q^\star} dF(y) - \tau \int_{q^\star}^{\infty} dF(y). \text{ i.e. } 0 = F_Y(\hat{q\star}) - \tau$$

i.e. $q^\star = Q(\tau)$. The empirical quantile for a sample $\boldsymbol{y}$, is solution of

$$\hat{q}(\tau) = \underset{q \in \mathbb{R}}{\text{argmin}} \left\{ (\tau - 1) \sum_{y_i < q} (y_i - q) + \tau \sum_{y_i \geq q} (y_i - q) \right\}$$

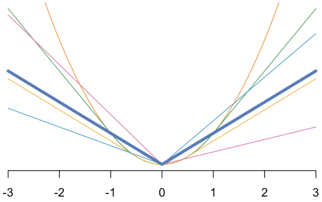
ℓ_2 and ℓ_1 loss function

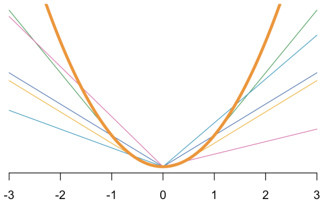
ℓ_2 (quadratic) loss

$$\ell(y,\widehat{y}) = (y-\widehat{y})^2 = c(y-\widehat{y}) \text{ where } c(x) = x^2$$

ℓ_1 (absolute) loss

$$\ell(y,\widehat{y}) = |y-\widehat{y}| = c(y-\widehat{y}) \text{ where } c(x) = |x|$$



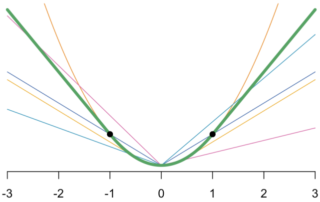


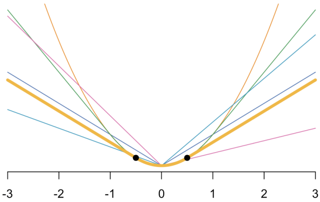
Huber loss function

Huber loss

$$\ell(y,\widehat{y}) = c(y-\widehat{y}) \text{ where } c(x) = \begin{cases} x^2 \text{ if } |x| < h \\ 2h|x| - h^2 \text{ if } |x| \geq h \end{cases}$$

Hybrid between quadratic ℓ_2 and ℓ_1, ℓ_2 for small error ℓ_1 for large ones. It is a convex and differentiable loss function



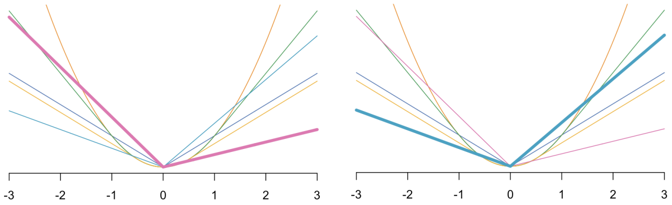


Quantile loss

Quantile (τ) loss

$$\ell(y,\widehat{y}) = c(y-\widehat{y}) \text{ where } c(x) = \begin{cases}(\tau-1)x \text{ if } x<0\\ \tau x \text{ if } x>0\end{cases}$$

asymmetric ℓ_1 loss function, or tilted absolute value function



0-1 and Hinge loss function

0-1 loss

$\ell(y, \widehat{y}) = \mathbf{1}(\widehat{y} \neq y)$

Note: ℓ is called a surrogate loss function if it is convex, and $\ell \geq \ell_{0-1}$.

Hinge loss

when $y \in \{0, 1\}$, and $\widehat{y} \in [0, 1]$ $\ell(y, \widehat{y}) = (1 - y\widehat{y})_+$

Logistic / Sigmoid / KL divergence / entropy loss function

Logistic loss

when $y \in \{0,1\}$, and $\widehat{y} \in [0,1]$

$$\ell(y,\widehat{y}) = -\big(y_i \log(\widehat{y}_i) + (1-y_i)\log(1-\widehat{y}_i)\big)$$

Note: can be extended to more than 2 categories, $\{0,1\}$
Note See also the softmax function, $\mathbb{R}^k \to \mathbb{R}^k$,

$$(x_1,\cdots,x_k) \mapsto \frac{1}{\sum_i e_i^x}\,(\hat{e}x_1,\cdots,e^{x_k})$$

Elicitable Measures

"*elicitable*" means "being a minimizer of a suitable expected score"
T is an elicitable function if there exits a scoring function $S:\mathbb{R}\times\mathbb{R}\to[0,\infty)$ such that, if $Y\sim F$,

$$T(Y)=\underset{x\in\mathbb{R}}{\text{argmin}}\left\{\int_{\mathbb{R}}S(x,y)dF(y)\right\}=\underset{x\in\mathbb{R}}{\text{argmin}}\left\{\mathbb{E}\big[S(x,Y)\big]\right\}$$

see Gneiting (2011).
Example: mean, $T(Y)=\mathbb{E}[Y]$ is elicited by $(x,y)=(x-y)^2$
Example: median, $T(Y)=\text{median}[Y]$ is elicited by $S(x,y)=|x-y|$
Example: quantile, $T(Y)=Q_Y(\tau)$ is elicited by $S(x,y)=\tau(y-x)_++(1-\tau)(y-x)_-$
Example: expectile, $T(Y)=E_Y(\tau)$ is elicited by $S(x,y)=\tau(y-x)_+^2+(1-\tau)(y-x)_-^2$