Dato un valore x, la $\Phi(x)$ standard function misura l'area da $-\infty$ a x

$$P(X < x) = \int_{-\infty}^{x} f(t)\,dt$$

La Survival function SF, misura l'area da x a $+\infty$

$$P(X > x) = \int_{x}^{+\infty} f(t)\,dt \longrightarrow \text{Nella teoria è la funzione } Q(x)$$

$$Q = 1 - \Phi(x)$$

+ importante

Nel criterio NP vogliamo ottimizzare $1-\beta \rightarrow \arg\min P(\hat{Y}=0 \mid Y=1)$

che è soggetta ad $P(\hat{Y}=1 \mid Y=0) \leq \alpha$

Con distribuzioni gaussiane con varianza uguali, il test della likelihood Ratio si limita a una soglia γ

Si sceglie $\hat{Y}=1$ se $x > \gamma \longrightarrow \dfrac{f(x \mid Y=1)}{f(x \mid Y=0)} \underset{0}{\overset{1}{\gtrless}} \gamma$

Dato che $x \mid Y=0 \sim N(0,1)$, la $P(\hat{Y}=1 \mid Y=0)$ è

$P(x > \gamma \mid Y=0) \longrightarrow$ Si usa la survive function $Q(\gamma) \leq \alpha$

Dalla teoria $P(\hat{Y}=1 \mid Y=0) = P(x > \gamma \mid Y=0) = P(N(\mu_0, \sigma^2) > \gamma) =$

$= P\left(\dfrac{N(\mu_0, \sigma^2) - \mu_0}{\sigma} > \dfrac{\gamma - \mu_0}{\sigma}\right) = Q\left(\dfrac{\gamma - \mu_0}{\sigma}\right) \rightarrow Q\left(\dfrac{\gamma - 0}{1}\right) = Q(\gamma)$

area da γ a $+\infty$

Se si impone $Q\left(\frac{\gamma - \mu_0}{\sigma}\right) = \alpha$ Dato che $Q\left(\frac{\gamma - \mu_0}{\sigma}\right) = Q(\gamma)$

$$\gamma = Q^{-1}(\alpha)$$

μ_0 = media di $\mathbb{P}(x \mid Y = 0)$
σ^2 varianza di $\mathbb{P}(x \mid Y = 0)$

Calcolo $\mathbb{P}(\hat{Y} = 1 \mid Y = 0)$ (Montecarlo) $\to$ Calcolo Empirico

$N°(Y=0)$: numero di campioni con $Y = 0$

$$\frac{1}{N°(Y=0)} \sum_{\alpha} \mathbb{I}(\hat{Y} = 1 \mid Y = 0)$$

$\hookrightarrow \mathbb{P}(\hat{Y} = 1 \mid Y = 0)$ falsi positivi $< 0,1$

Requisito Rispettato

Su quelli effettivamente negativi, vedi quanti ne hai classificati correttamente

$\mathbb{P}(\hat{Y} = 0 \mid Y = 1)$ 0,58 il 58% delle osservazioni viene ignorato o sbagliato per far rientrare $\mathbb{P}(\hat{y} = 1 \mid y = 0) \leq \alpha$

ESERCIZIO SGD

$x \in \mathbb{R}^2$ $y \in \{-1, 1\}$

Si hanno che le feature $x_1 \mid Y = 1$ e $x_2 \mid Y = -1$ $\mu = 0,9$

$x_1 \mid Y = -1$ e $x_2 \mid Y = 1$ $\mu = -0,85$

Definire il calcolo del gradiente

Definiamo una statistica di decisione $S_\beta(x) \triangleq \ln \frac{P(1|x)}{P(-1|x)} = \ln \frac{P(1|x)}{1-P(1|x)}$

$P(1|x) = \frac{1}{1+e^{-S_\beta(x)}}$ ← Soft form della posterior

La loss function nel caso di Regressione Logistica, in cui la statistica assume una forma lineare è

$Q_\beta(x,y) = \ln(1+e^{-yx^T\beta})$

Applicando la teoria del gradiente: $\nabla J(\beta)$ $\frac{\delta J(\beta)}{\delta(\beta_i)}$

$\nabla_\beta Q = \frac{1}{1+e^{-yx^T\beta}}\left(e^{-yx^T\beta}\cdot(-yx)\right) = \frac{-yx}{1+e^{yx^T\beta}}$

$x^T = 1 \cdot n$ poteva permettere il prodotto con $\beta(n \cdot 1)$

Se si ha $\nabla_\beta(x^T\beta) = x$

Punto e

Classificazione MAP: $P(Y=y|X=x) = \frac{\ell(X=x|Y=y)\,\pi(y)}{\sum \pi(y')\,\ell(x|y')}$

con $y' \in \{y_1, y_2\}$ è uguale ad 1

$\frac{P(Y=1|X=x)}{P(Y=-1|X=x)} = \frac{\ell(x|Y=1)\,\pi(Y=1)}{\ell(x|Y=-1)\,\pi(Y=-1)} \underset{-1}{\overset{1}{\gtrless}} \phi$

$\frac{\ell(x|1)}{\ell(x|-1)} \underset{-1}{\overset{1}{\gtrless}} \cancel{\frac{\pi(1)}{\pi(-1)}}$ le classi sono ugualmente distribuite $\pi(1)=\pi(-1)$

$$p(x|Y=1) \longrightarrow p(x_1|Y=1)\,p(x_2|Y=1)$$

$$p(x_1|1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_1+0{,}85)^2}{2}\right)$$

$$p(x_2|1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_2-90)^2}{2}\right)$$

$$p(x|Y=1) = \frac{1}{2\pi} \exp\left(-\frac{(x_1+0{,}85)^2+(x_2-0{,}90)^2}{2}\right)$$

$$p(x|Y=-1) = p(x_1|-1)\,p(x_2|-1) = \frac{1}{2\pi} \exp\left(-\frac{(x_1-0{,}9)^2+(x_2+0{,}85)^2}{2}\right)^2$$

log likelihood Ratio $p(x|1) - \log(x|-1) \geq \phi$ allora x è $Y=1$

$$-\frac{1}{2}\left((x_1+0{,}85)^2+(x_2-0{,}9)^2\right) + \frac{1}{2}\left((x_1-0{,}9)^2+(x_2+0{,}85^2)\right)$$

Sviluppando i quadrati

$x_1^2 + 1{,}7x_1 + 0{,}72$

$x_2^2 - 1{,}8x_2 + 0{,}81$

$x_1^2 - 1{,}8\,x_1 + 0{,}81$

$x_2^2 + 1{,}7x_2 + 0{,}72$

Andando a sostituire

$-1{,}75x_1 + 1{,}75x_2 \geq \phi$ $\qquad \cancel{1{,}7}5x_2 \geq \cancel{1{,}7}5x_1 \qquad x_1 = x_2$

Retta passante per l'origine $\beta_1 = -1{,}75$ $\beta_2 = 1{,}75$ $\beta_\emptyset = \emptyset$