Application of GLM Advancements to Non-Life Insurance Pricing

Leonardo Stincone

Università degli Studi di Trieste

11 Maggio 2021

Indice

Indice

Contratto di Assicurazione, Art. 1882, Codice Civile Italiano

L'assicurazione è il contratto col quale l'**assicuratore**, verso il pagamento di un **premio**, si obbliga a rivalere l'**assicurato**, entro i limiti convenuti,

1. del **danno** ad esso prodotto da un **sinistro**,
2. ovvero a pagare un **capitale** o una **rendita** al verificarsi di un **evento** attinente alla **vita umana**.

Da un punto di vista matematico

	t_1	τ_1	τ_2	$\dots$	τ_{N-1}	τ_N	t_2	t
Assicurato	$-P$							
Assicuratore	P	$-Z_1$	$-Z_2$		$-Z_{N-1}$	$-Z_N$		

Distribuzione composta

Assumiamo che:

1. $\forall n > 0,\ Z_1|N = n,\ Z_2|N = n,\ \dots,\ Z_n|N = n$ siano i.i.d.;
2. la distribuzione di $Z_i|N = n,\ i \leq n$ non dipenda da n.

Sotto queste ipotesi diciamo che:

$$S = \begin{cases} 0 & \text{if } N = 0 \\ \sum_{i=1}^{N} Z_i & \text{if } N > 0 \end{cases}$$

ha distribuzione composta.

Proprietà

$$\begin{tikzcd} F_N \arrow[dr] & & F_Z \arrow[dl] \\ & F_S & \end{tikzcd}$$

$$E(S) = E(N)E(Z)$$

	t_1	τ_1	τ_2	$\dots$	τ_{N-1}	τ_N	t_2	t
Assicurato	$-P$							
Assicuratore	P	$-Z_1$	$-Z_2$		$-Z_{N-1}$	$-Z_N$		

Distribuzione composta

Assumiamo che:

1. $\forall n > 0,\ Z_1|N = n,\ Z_2|N = n,\ \dots,\ Z_n|N = n$ siano i.i.d.;
2. la distribuzione di $Z_i|N = n,\ i \leq n$ non dipenda da n.

Sotto queste ipotesi diciamo che:

$$S = \begin{cases} 0 & \text{if } N = 0 \\ \sum_{i=1}^{N} Z_i & \text{if } N > 0 \end{cases}$$

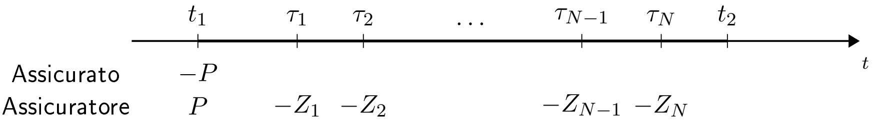
ha distribuzione composta.

Proprietà

$$\begin{tikzcd} F_N \arrow[dr] & & F_Z \arrow[dl] \\ & F_S & \end{tikzcd}$$

$$E(S) = E(N)E(Z)$$

Da un punto di vista matematico



Distribuzione composta

Assumiamo che:

1. $\forall n > 0,\ Z_1|N = n,\ Z_2|N = n,\ \dots,\ Z_n|N = n$ siano i.i.d.;
2. la distribuzione di $Z_i|N = n,\ i \leq n$ non dipenda da n.

Sotto queste ipotesi diciamo che:

$$S = \begin{cases} 0 & \text{if } N = 0 \\ \sum_{i=1}^{N} Z_i & \text{if } N > 0 \end{cases}$$

ha distribuzione composta.

Proprietà

$$\begin{tikzcd} F_N \arrow[dr] & & F_Z \arrow[dl] \\ & F_S & \end{tikzcd}$$

$$E(S) = E(N)E(Z)$$

Variabili Esplicative e Personalizzazione

Variabili esplicative

Possibili variabili esplicative per il pricing delle assicurazioni motor:

- Informazioni sul veicolo assicurato;
- Informazioni generiche sull'assicurato;
- Informazioni assicurative sull'assicurato;
- Opzioni della polizza assicurativa;
- Informazioni sull'assicurato in quanto cliente;
- Dati telematici.

Queste variabili possono essere codificate come un vettore di numeri reali:

$$\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip}) \in \mathcal{X} \subseteq \mathbb{R}^p$$

Regola di Pricing

Una *Regola di Pricing* è una funzione $f(\cdot)$ che da una $\boldsymbol{x}_i \in \mathcal{X}$ restituisce un prezzo P_i:

$$\begin{aligned} f: \quad \mathcal{X} &\longrightarrow R_+ \\ \boldsymbol{x}_i &\longmapsto P_i \end{aligned}$$

Modellare una variabile risposta

Modellare una variabile risposta Y_i significa stimare una funzione $r(\cdot)$ che da una $\boldsymbol{x}_i \in \mathcal{X}$ restituisce la distribuzione di Y_i o alcuni suoi momenti:

$$\begin{aligned} r: \quad \mathcal{X} &\longrightarrow \mathcal{C} \\ \boldsymbol{x}_i &\longmapsto F_{Y_i}, E(Y_i), Var(Y_i) \end{aligned}$$

Variabili esplicative

Possibili variabili esplicative per il pricing delle assicurazioni motor:

- Informazioni sul veicolo assicurato;
- Informazioni generiche sull'assicurato;
- Informazioni assicurative sull'assicurato;
- Opzioni della polizza assicurativa;
- Informazioni sull'assicurato in quanto cliente;
- Dati telematici.

Queste variabili possono essere codificate come un vettore di numeri reali:

$$\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip}) \in \mathcal{X} \subseteq \mathbb{R}^p$$

Regola di Pricing

Una *Regola di Pricing* è una funzione $f(\cdot)$ che da una $\boldsymbol{x}_i \in \mathcal{X}$ restituisce un prezzo P_i:

$$\begin{array}{rccc} f: & \mathcal{X} & \longrightarrow & R_+ \\ & \boldsymbol{x}_i & \longmapsto & P_i \end{array}$$

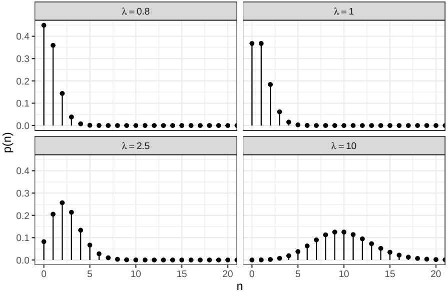
Modellare una variabile risposta

Modellare una variabile risposta Y_i significa stimare una funzione $r(\cdot)$ che da una $\boldsymbol{x}_i \in \mathcal{X}$ restituisce la distribuzione di Y_i o alcuni suoi momenti:

$$\begin{array}{rccc} r: & \mathcal{X} & \longrightarrow & \mathcal{C} \\ & \boldsymbol{x}_i & \longmapsto & F_{Y_i}, E(Y_i), Var(Y_i) \end{array}$$

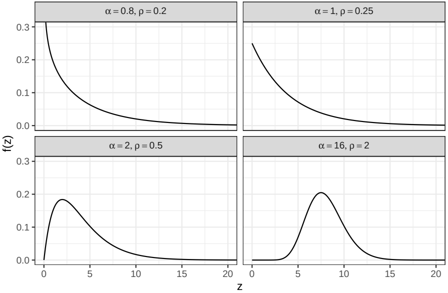
Distribuzione di Poisson

$$p_N(n) = P(N = n) = e^{-\lambda} \frac{\lambda^n}{n!}, \quad \lambda > 0$$



Distribuzione Gamma

$$f_Z(z) = \frac{\rho^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\rho z}, \quad \alpha > 0, \ \rho > 0$$



Definizione di Premio

$$P_i^{(\mathrm{risk})} = E(S_i)$$
$$P_i^{(\mathrm{tech})} = E(S_i) + \mathrm{Expenses}_i$$

Altri Caricamenti
Vincoli Normativi
Commercializzazioni

$$P_i^{(\mathrm{tariff})}$$
$$P_i^{(\mathrm{offer})} = P_i^{(\mathrm{tariff})} - \mathrm{Discount}_i$$

Ottimizzazione del Prezzo

Si basa su

1. Pricing Tecnico
2. Aspettativa del Cliente
   - New Business: *Probabilità di Conversion*
   - Rinnovi: *Probabilità di Retention*
3. Strategia di Business
   - *Lifetime value*
   - Profitti/crescita

Definizione di Premio

$$P_i^{(\text{risk})} = E(S_i)$$
$$P_i^{(\text{tech})} = E(S_i) + \text{Expenses}_i$$

Altri Caricamenti
Vincoli Normativi
Commercializzazioni

$$P_i^{(\text{tariff})}$$
$$P_i^{(\text{offer})} = P_i^{(\text{tariff})} - \text{Discount}_i$$

Ottimizzazione del Prezzo

Si basa su

1. Pricing Tecnico
2. Aspettativa del Cliente
   - New Business: *Probabilità di Conversion*
   - Rinnovi: *Probabilità di Retention*
3. Strategia di Business
   - *Lifetime value*
   - Profitti/crescita

Indice

Modelli Lineari Generalizzati (GLM)

Dato $\mathcal{D} = \{(\boldsymbol{x}_1, \omega_1, y_1), \ldots, (\boldsymbol{x}_n, \omega_n, y_n)\}$
con $\boldsymbol{y} = (y_1, \ldots, y_n)^t$ realizzazione di $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^t$.

Assumiamo che:

1. $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^t$ siano indipendenti con distribuzione appartenente a una stessa famiglia esponenziale lineare:

$$f(y_i; \theta_i, \phi, \omega_i) = \exp\left\{\frac{\omega_i}{\phi}[y_i\theta_i - b(\theta_i)]\right\} c(y_i, \phi, \omega_i), \quad y_i \in \mathcal{Y} \subseteq \mathbb{R}$$

2. $\boldsymbol{x}_i = (1, x_{i1}, \ldots, x_{ip})^t$ agisca su Y_i tramite il predittore lineare η_i

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}$$

3. η_i sia legato a $\mu_i = E(Y_i)$ tramite la funzione legame $g(\cdot)$

$$g(\mu_i) = \eta_i = \boldsymbol{x}_i^t \boldsymbol{\beta}$$

Stima di un GLM

Stima di massima verosimiglianza

Data la funzione di verosimiglianza

$$
\begin{array}{rccc}
L: & \mathbb{R}^{p+1} \times \Lambda & \longrightarrow & [0, +\infty[\\
 & (\boldsymbol{\beta}, \phi) & \longmapsto & f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta}, \phi)
\end{array}
$$

La stima di massima verosimiglianza è:

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta} \in \mathbb{R}^{p+1}} L(\boldsymbol{\beta}, \phi; \boldsymbol{y})$$

Devianza

La devianza è

$$D(\hat{\boldsymbol{\beta}}, \boldsymbol{y}) = -2\phi \left(\ell\left(\hat{\boldsymbol{\beta}}, \phi; \boldsymbol{y}\right) - \ell_S\left(\boldsymbol{\beta}^*, \phi; \boldsymbol{y}\right) \right)$$

dove $\ell\left(\hat{\boldsymbol{\beta}}, \phi; \boldsymbol{y}\right) = \log L\left(\hat{\boldsymbol{\beta}}, \phi; \boldsymbol{y}\right)$
e $\boldsymbol{\beta}^*$ sono i parametri del modello saturo.

La stima di massima verosimiglianza può essere ottenuta come:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^{p+1}} D(\boldsymbol{\beta}, \boldsymbol{y})$$

Stima di massima verosimiglianza

Data la funzione di verosimiglianza

$$\begin{array}{rccc} L: & \mathbb{R}^{p+1} \times \Lambda & \longrightarrow & [0, +\infty[\\ & (\boldsymbol{\beta}, \phi) & \longmapsto & f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta}, \phi) \end{array}$$

La stima di massima verosimiglianza è:

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta} \in \mathbb{R}^{p+1}} L(\boldsymbol{\beta}, \phi; \boldsymbol{y})$$

Devianza

La devianza è

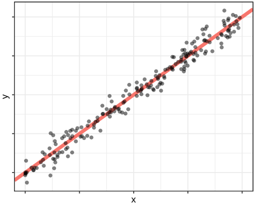
$$D(\hat{\boldsymbol{\beta}}, \boldsymbol{y}) = -2\phi \left(\ell\left(\hat{\boldsymbol{\beta}}, \phi; \boldsymbol{y}\right) - \ell_S(\boldsymbol{\beta}^*, \phi; \boldsymbol{y}) \right)$$

dove $\ell\left(\hat{\boldsymbol{\beta}}, \phi; \boldsymbol{y}\right) = \log L\left(\hat{\boldsymbol{\beta}}, \phi; \boldsymbol{y}\right)$
e $\boldsymbol{\beta}^*$ sono i parametri del modello saturo.

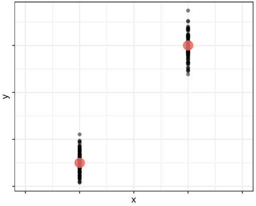
La stima di massima verosimiglianza può essere ottenuta come:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^{p+1}} D(\boldsymbol{\beta}, \boldsymbol{y})$$

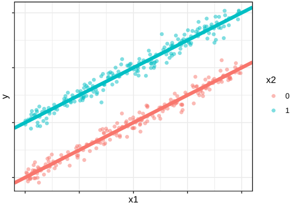
Quantitativa



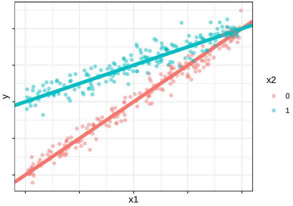
Qualitativa



Quantitativa e qualitativa
senza interazione

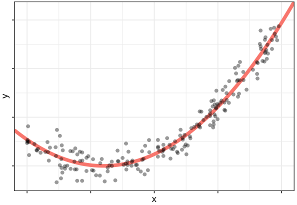


Quantitativa e qualitativa
con interazione

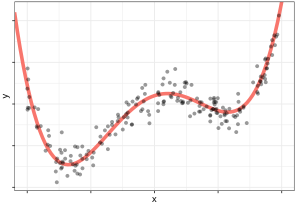


Variabili quantitative ed effetti non lineari

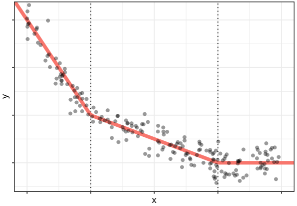
Polinomiale di grado 2



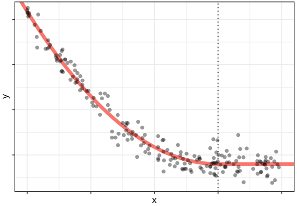
Polinomiale di grado 4



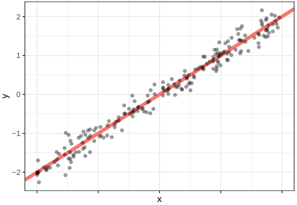
Piece-wise lineare



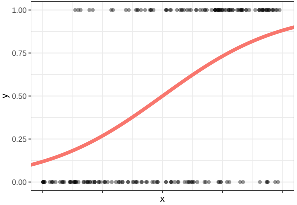
Piece-wise polinomiale di grado 2



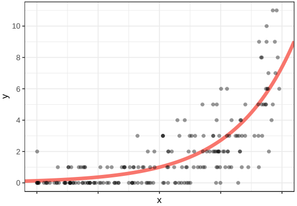
Normale - identità



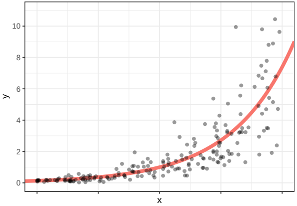
Binomiale - logit



Poisson - log

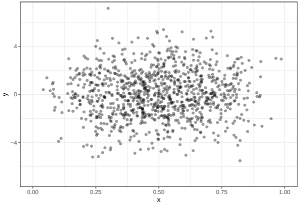


Gamma - log

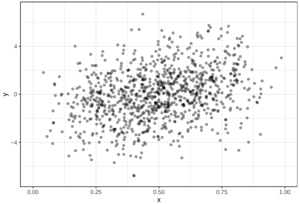


Grafici per visualizzare l'effetto delle variabili

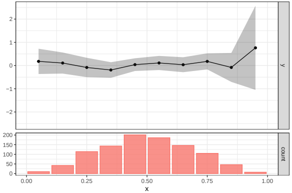
Nessun effetto - non raggruppati



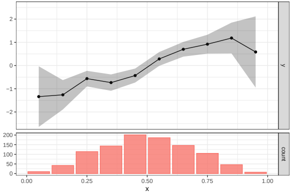
Effetto positivo - non raggruppati



Nessun effetto - raggruppati



Effetto positivo - raggruppati



Criteri per la selezione delle variabili

- Visualizzazione
- Test di verifica di ipotesi

$$\begin{cases} H_0: & \beta_{j_k} = 0 \ \forall k \in \{1, 2, \ldots, s\} \\ H_1: & \exists k : \beta_{j_k} \neq 0 \end{cases}$$

- Criteri di informazione

$$AIC = -2\ell(\boldsymbol{\beta}) + 2(p+1)$$
$$BIC = -2\ell(\boldsymbol{\beta}) + \log(n)(p+1)$$

- Divisione del dataset tra training set e test set
- Cross validation

$\Longrightarrow$ Algoritmi stepwise

Modelli Additivi Generalizzati (GAM)

1. Variabile risposta $\boldsymbol{Y}$ come GLM;
2. Predittore lineare

$$\eta_i = \boldsymbol{x}_i^t \boldsymbol{\beta} + \sum_{l=1}^{q} f_l(z_{i,l}), \quad i \in \{1, 2, \ldots, n\}$$

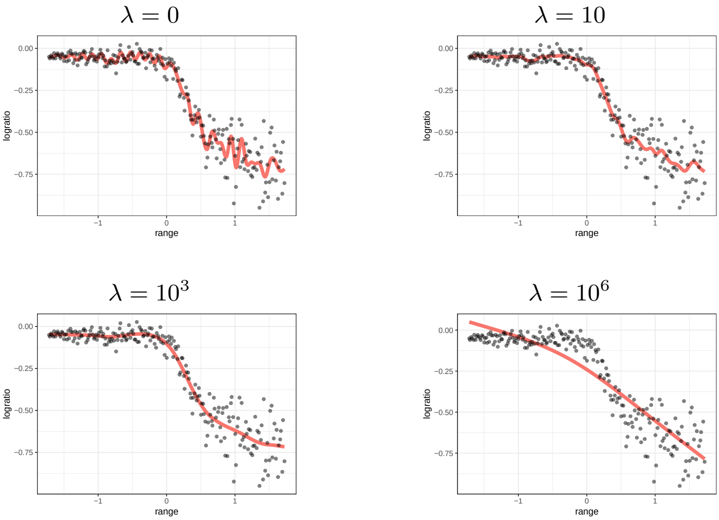
   con $f_l(\cdot)$ spline cubica;
3. Funzione legame $g(\cdot)$ come GLM.

Stima di Massima Verosimiglianza con Penalizzazione

$$\hat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \left\{ D(\boldsymbol{f}, \boldsymbol{y}) + \sum_{l=1}^{q} \lambda_l \int_{a_l}^{b_l} \left(f_l''(x_l)\right)^2 dx \right\}$$

con $\lambda_1, \lambda_2, \ldots, \lambda_q$ iperparametri di smoothing.

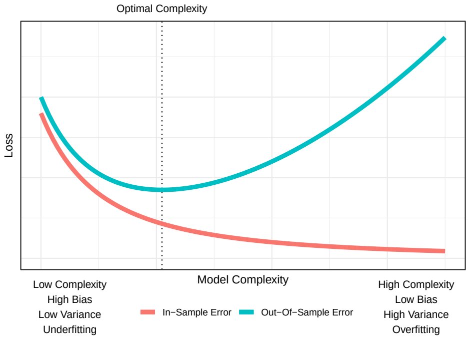
GAM: esempio



$\lambda = 0$
$\lambda = 10$
$\lambda = 10^3$
$\lambda = 10^6$
0.00
−0.25
−0.50
−0.75
logratio
−1
0
1
range

Scomposizione dello scarto quadratico medio (MSE)

$$MSE\left(\tilde{\beta}_j\right) \overset{\text{def}}{=} E\left(\left(\tilde{\beta}_j - \beta_j\right)^2\right) = \underbrace{\left(E(\tilde{\beta}_j) - \beta_j\right)^2}_{\text{Bias}^2} + \underbrace{Var\left(\tilde{\beta}_j\right)}_{\text{Variance}}$$



Stima di Massima Verosimiglianza con Penalizzazione

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}}{\arg\min} \left\{ D(\boldsymbol{\beta}, \boldsymbol{y}) + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}$$

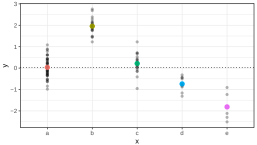
con

- $\lambda \geq 0$ iperparametro di penalizzazione

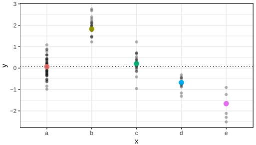
Modello sottostante:GLM

Regressione Ridge: esempio

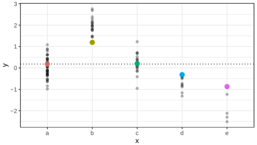
$\lambda = 0$



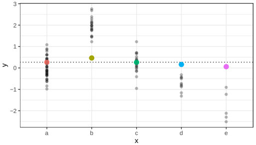
$\lambda = 0.1$

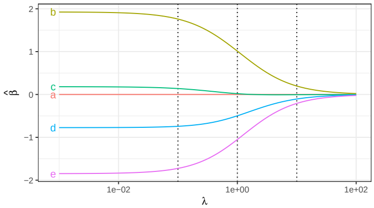


$\lambda = 1$



$\lambda = 10$





Stima di Massima Verosimiglianza con Penalizzazione

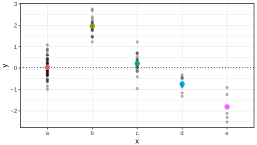
$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}}{\arg\min} \left\{ D(\boldsymbol{\beta}, \boldsymbol{y}) + \lambda \sum_{j=1}^{p} |\beta_j| \right\}$$

con

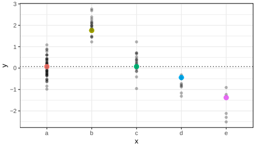
- $\lambda \geq 0$ iperparametro di penalizzazione

Modello sottostante:GLM

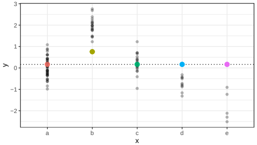
$\lambda = 0$



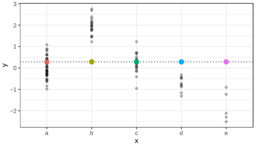
$\lambda = 0.1$

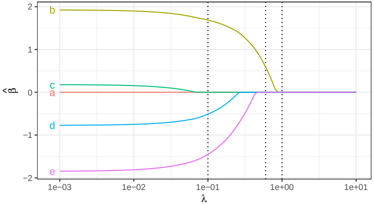


$\lambda = 1$



$\lambda = 10$





Stima di Massima Verosimiglianza con Penalizzazione

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}}{\arg\min} \left\{ D(\boldsymbol{\beta}, \boldsymbol{y}) + \lambda \sum_{j=1}^{p} \left(\alpha |\beta_j| + (1-\alpha)|\beta_j|^2 \right) \right\}$$

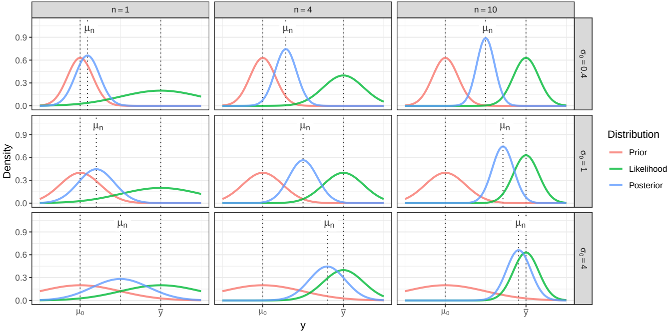
dove

- $\lambda \geq 0$ iperparametro di penalizzazione
- $\alpha \in [0, 1]$ iperparametro che determina il peso della penalizzazione LASSO
  - $\alpha = 0 \implies$ Regressione Ridge
  - $\alpha = 1 \implies$ Regressione LASSO

Modello sottostante:GLM

Teorema di Bayes

$$\pi(\theta|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\theta)\pi(\theta)}{p(\boldsymbol{y})}$$



Stimatori Bayesiani per i GLM

Stima di Massima Verosimiglianza

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}^{ML} &= \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}}{\arg\max}\, L\left(\boldsymbol{\beta}, \phi \mid \boldsymbol{y}\right) \\
&= \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}}{\arg\max}\, \ell\left(\boldsymbol{\beta}, \phi \mid \boldsymbol{y}\right) \\
&= \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}}{\arg\min}\, D\left(\boldsymbol{\beta}, \boldsymbol{y}\right)
\end{aligned}
$$

Stima di Massimo a Posteriori

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}^{MAP} &= \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}}{\arg\max}\, \pi\left(\boldsymbol{\beta}, \phi \mid \boldsymbol{y}\right) \\
&= \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}}{\arg\max}\, \left\{L\left(\boldsymbol{\beta}, \phi \mid \boldsymbol{y}\right) \pi(\boldsymbol{\beta})\right\} \\
&= \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}}{\arg\max}\, \left\{\ell\left(\boldsymbol{\beta}, \phi \mid \boldsymbol{y}\right) + \log\left(\pi(\boldsymbol{\beta})\right)\right\} \\
&= \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}}{\arg\min}\, \left\{D(\boldsymbol{\beta}, \boldsymbol{y}) - 2\phi \log\left(\pi(\boldsymbol{\beta})\right)\right\}
\end{aligned}
$$

Stima di Massima Verosimiglianza

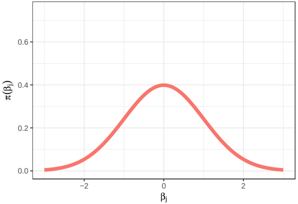
$$
\begin{aligned}
\hat{\boldsymbol{\beta}}^{ML} &= \underset{\boldsymbol{\beta}\in\mathbb{R}^{p+1}}{\arg\max}\, L\left(\boldsymbol{\beta},\phi \mid \boldsymbol{y}\right) \\
&= \underset{\boldsymbol{\beta}\in\mathbb{R}^{p+1}}{\arg\max}\, \ell\left(\boldsymbol{\beta},\phi \mid \boldsymbol{y}\right) \\
&= \underset{\boldsymbol{\beta}\in\mathbb{R}^{p+1}}{\arg\min}\, D\left(\boldsymbol{\beta},\boldsymbol{y}\right)
\end{aligned}
$$

Stima di Massimo a Posteriori

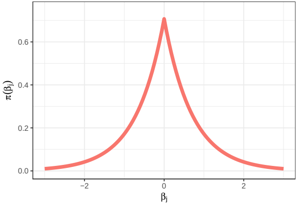
$$
\begin{aligned}
\hat{\boldsymbol{\beta}}^{MAP} &= \underset{\boldsymbol{\beta}\in\mathbb{R}^{p+1}}{\arg\max}\, \pi\left(\boldsymbol{\beta},\phi \mid \boldsymbol{y}\right) \\
&= \underset{\boldsymbol{\beta}\in\mathbb{R}^{p+1}}{\arg\max}\, \left\{L\left(\boldsymbol{\beta},\phi \mid \boldsymbol{y}\right)\pi(\boldsymbol{\beta})\right\} \\
&= \underset{\boldsymbol{\beta}\in\mathbb{R}^{p+1}}{\arg\max}\, \left\{\ell\left(\boldsymbol{\beta},\phi \mid \boldsymbol{y}\right) + \log\left(\pi(\boldsymbol{\beta})\right)\right\} \\
&= \underset{\boldsymbol{\beta}\in\mathbb{R}^{p+1}}{\arg\min}\, \left\{D(\boldsymbol{\beta},\boldsymbol{y}) - 2\phi\log\left(\pi(\boldsymbol{\beta})\right)\right\}
\end{aligned}
$$

Regressione Ridge e LASSO come Stimatori Bayesiani

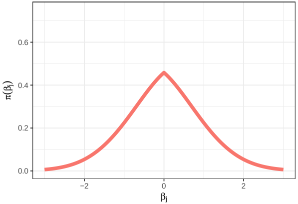
Distribuzione Normale
$\Rightarrow$ Regressione Ridge



Distribuzione di Laplace
$\Rightarrow$ Regressione LASSO



Distribuzion intermedia
$\Rightarrow$ Elastic Net



Considerazioni sugli Stimatori Bayesiani

Altre distribuzioni a priori

- Diverse varianze a priori

$$\beta_j \sim \mathcal{N}(0, \sigma_j^2)$$

- Diverse medie a priori

$$\beta_j \sim \mathcal{N}(\beta_{j0}, \sigma_j^2)$$

- Altre distribuzioni a priori

$$\pi(\beta_j) = \begin{cases} \frac{\sqrt{2}}{\sqrt{\pi}\sigma} e^{-\frac{1}{2\sigma^2}\beta_j^2} & \text{if } \beta_j \geq 0 \\ 0 & \text{altrimenti} \end{cases}$$

Vantaggi stimatori Bayesiani

- Introduzione informazione esterna ai dati con una robusta metodologia statistica
- Rimpiazzamento degli offset
  Scelgo σ_j^2 tale che $\hat{\beta}^{MAP} = \hat{\beta}^{\text{offset}}$
  - Ho accortezza di quanto è forte la correzzione applicata
  - Se cambio qualche altro parametro e rifitto il modello, in automatico $\hat{\beta}^{MAP}$ viene ristimato

Altre distribuzioni a priori

- Diverse varianze a priori

$$\beta_j \sim \mathcal{N}(0, \sigma_j^2)$$

- Diverse medie a priori

$$\beta_j \sim \mathcal{N}(\beta_{j0}, \sigma_j^2)$$

- Altre distribuzioni a priori

$$\pi(\beta_j) = \begin{cases} \frac{\sqrt{2}}{\sqrt{\pi}\sigma} e^{-\frac{1}{2\sigma^2}\beta_j^2} & \text{if } \beta_j \geq 0 \\ 0 & \text{altrimenti} \end{cases}$$

Vantaggi stimatori Bayesiani

- Introduzione informazione esterna ai dati con una robusta metodologia statistica
- Rimpiazzamento degli offset
  Scelgo σ_j^2 tale che $\hat{\boldsymbol{\beta}}^{MAP} = \hat{\boldsymbol{\beta}}^{\text{offset}}$
  - Ho accortezza di quanto è forte la correzione applicata
  - Se cambio qualche altro parametro e rifitto il modello, in automatico $\hat{\boldsymbol{\beta}}^{MAP}$ viene ristimato

Modelli di Machine Learning

- Gradient Boosting Machine (GBM)
- Random Forest (RF)
- Neural Network (NN)
- Altri . . .

Caratteristiche

- Funzione di regressione con minime assunzioni

$$E(Y_i) = \boldsymbol{f}(x_{i1}, \ldots, x_{ip})$$

- Sofisticati algoritmi per prevenire l'overfitting

Modelli di Machine Learning

- Gradient Boosting Machine (GBM)
- Random Forest (RF)
- Neural Network (NN)
- Altri ...

Caratteristiche

- Funzione di regressione con minime assunzioni

$$E(Y_i) = \boldsymbol{f}(x_{i1}, \ldots, x_{ip})$$

- Sofisticati algoritmi per prevenire l'overfitting

Confronto tra i modelli

	GLM Classici	GBM/RF/NN	GLM Advancements
Interpretabilità	★★★	★☆☆	★★★
Controllo delle variabili	★★★	★☆☆	★★★
Utilizzo di informazioni esterne	★★☆	★☆☆	★★★
Automazione e scalabilità	★☆☆	★★★	★★☆
Flessibilità	★☆☆	★★★	★★☆

L'importanza del Controllo delle Variabili nel Pricing

Il pricing si basa su

- Osservazioni sul portafoglio passato
- Assunzioni sul portafoglio futuro

Necessità tecniche
per il controllo delle variabili

- Dati non rappresentativi del portafoglio futuro
- Stime con alta varianza su certi cluster

Necessità commerciali
per il controllo delle variabili

- Vincoli normativi
- Pricing opzioni
- Aspettative del cliente
- Strategia di business

Il pricing si basa su

- Osservazioni sul portafoglio passato
- Assunzioni sul portafoglio futuro

Necessità tecniche
per il controllo delle variabili

- Dati non rappresentativi del portafoglio futuro
- Stime con alta varianza su certi cluster

Necessità commerciali
per il controllo delle variabili

- Vincoli normativi
- Pricing opzioni
- Aspettative del cliente
- Strategia di business

Origine del Dataset

Portafogio RCA costituito da polizze di una provincia italiana nel periodo 2014-2019

Set	Osservazioni	Esposizione (rischi anno)	Assicurati	Esposizione per Assicurato	Numero Sinistri	Frequenza Sinistri
Train	227 226	107 998.4	27 346	3.95	4 823	0.045
Test	56 603	26 806.3	6 824	3.93	1 131	0.042
Tot	283 829	134 804.7	34 170	3.95	5 954	0.044

Descrizione	**Numero di variabili per categoria**
Informazioni sul veicolo assicurato	12
Informazioni generiche sull'assicurato	14
Informazioni assicurative sull'assicurato	9
Opzioni della polizza assicurativa	11
Informazioni sull'assicurato in quanto cliente	2
Dati telematici	4
Totale	**52**

Modelli considerati

Id	Model
Mod1	GLM Tot
Mod2	Elastic Net Tot
Mod3	Ridge Tot
Mod4	GLM AIC
Mod5	Elastic Net AIC
Mod6	GAM AIC
Mod7	GBM Tot

Valutazione

Metrica di confronto:
Devianza della distribuzione di Poisson calcolata sul test set

$$D(\hat{\boldsymbol{\beta}}, \boldsymbol{y}) = 2 \sum_{i=1}^{n} \left\{ y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i) \right\}$$

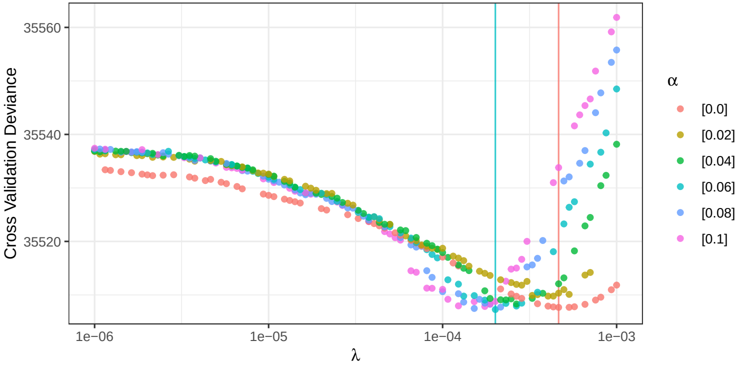
Modelli considerati

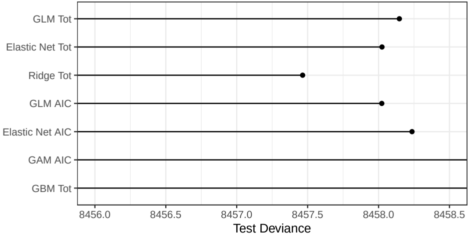
Id	Model
Mod1	GLM Tot
Mod2	Elastic Net Tot
Mod3	Ridge Tot
Mod4	GLM AIC
Mod5	Elastic Net AIC
Mod6	GAM AIC
Mod7	GBM Tot

Valutazione

Metrica di confronto:
Devianza della distribuzione di Poisson calcolata sul test set

$$D(\hat{\boldsymbol{\beta}}, \boldsymbol{y}) = 2 \sum_{i=1}^{n} \left\{ y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i) \right\}$$


Cross Validation Deviance
35560
35540
35520
1e−06
1e−05
1e−04
1e−03
λ
α
[0.0]
[0.02]
[0.04]
[0.06]
[0.08]
[0.1]



Id	Model	Test Deviance	Time	α	λ
Mod1	GLM Tot	8 458.147	2.7s	0	0
Mod2	Elastic Net Tot	8 458.024	1h 30m	0.06	2.01e-04
Mod3	**Ridge Tot**	**8 457.465**	1h 30m	0	4.64e-04
Mod4	GLM AIC	8 458.023	7h 27m	0	0
Mod5	Elastic Net AIC	8 458.236	8h 54m	0	1.63e-05
Mod6	GAM AIC	9 728.570	7h 45m	0	0
Mod7	GBM Tot	8 504.178	2h 30m		

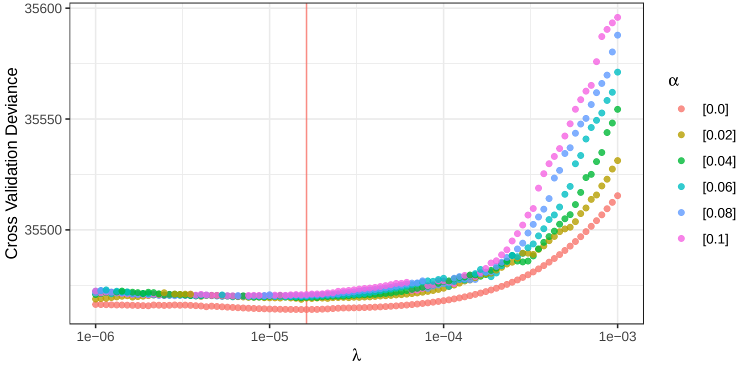
Possibili Sviluppi

- Utilizzare diverse distribuzioni a priori per i diversi β_j
- Considerare interazioni
- Ulteriore esplorazione sui GAM
- Modelli geografici
- Implementare i modelli su dataset più grandi con adeguate implementazioni informatiche

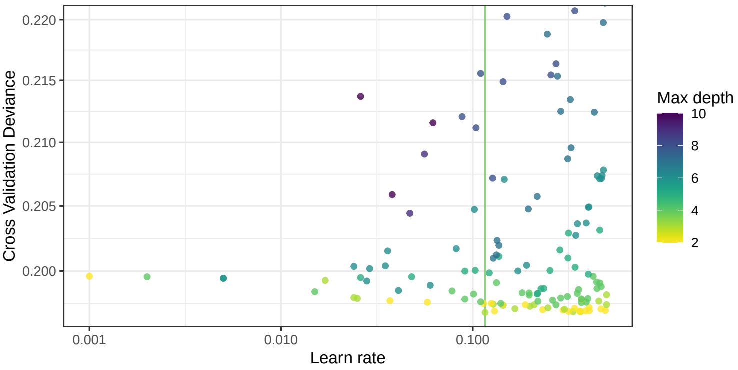
Grazie per l'attenzione

Backup

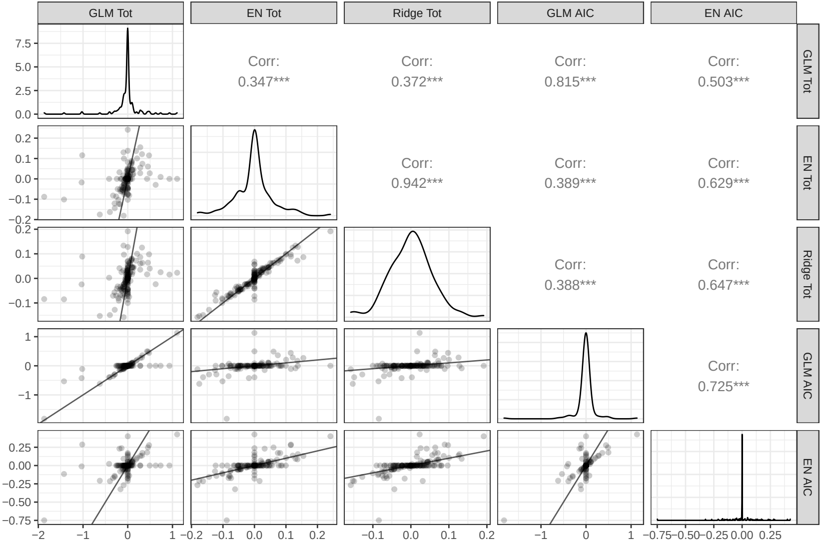
Elastic Net AIC: Tuning degli Iperparametri



Elastic Net AIC: Tuning degli Iperparametri



Confronto tra i coefficienti



Confronto tra i coefficienti: outlier

model	**value**
GLM Tot	-6.447
EN Tot	0.000
Ridge Tot	-0.003
GLM AIC	-6.574
EN AIC	-0.095

Coefficienti azzerati

GLM Tot	Elastic Net Tot	GLM AIC	n	
$\neq 0$	$\neq 0$	$\neq 0$	51	121
$\neq 0$	$\neq 0$	0	48	
$\neq 0$	0	$\neq 0$	6	
$\neq 0$	0	0	16	
0	$\neq 0$	0	23	38
0	0	0	15	