



Ambrogio Lorenzetti, Effects of Good Government, 1337-40, Palazzo Pubblico, Siena

Alef

Claim watching e Chain Ladder updating

Luca D'Agostino

Milano, 27 Maggio 2019

Indice degli argomenti

■ Letteratura di riferimento

1. Machine learning

1.1 ClaimWatching

1.2 Lo schema generale di “regressione”

1.3 Previsione di eventi a un anno

1.4 Previsione di eventi multiperiodale

1.5 Previsione di costi

1.6 Previsioni di *cost development* del singolo sinistro

1.7 Estensioni e potenziamenti del modello

2. Improved Chain Ladder

2.1 Combined Actuarial Neural Net

2.2 Il modello ODP classico

2.3 Struttura algoritmica della rete neurale

2.4 Esempi

Letteratura di riferimento

● Uso di *micro-level data* per la stima della riserva sinistri

[[Granular Models](#)]

- ⊕ Taylor G., McGuire G., Sullivan J. (2008). *Individual claim loss reserving conditioned by case estimates*. *Annals of Actuarial Science* 3/1-2, 215-256.
- ⊕ Pigeon M., Antonio K., Denuit M. (2013) *Individual loss reserving with the multivariate skew normal framework*. *ASTIN Bulletin* 43/3, 399-428.
- ⊕ Antonio K., Plat R. (2014) *Micro-level stochastic loss reserving for general insurance*. *Scandinavian Actuarial Journal* 2014/7, 649-669.
- ⊕ Martínez-Miranda M.D., Nielsen J.P., Verrall R.J., Wüthrich M.V. (2015). *The link between classical reserving and granular reserving through double chain ladder and its extensions*. *Scandinavian Actuarial Journal* 2015/5, 383-405.
- ⊕ Hiabu M., Margraff C., Martínez-Miranda M.D., Nielsen J.P. (2016). *The link between classical reserving and granular reserving through double chain ladder and its extensions*. *British Actuarial Journal* 21/1, 97-116.
- ⊕ Lopez O. (2018). *A censored copula model for micro-level claim reserving*. HAL Id: hal-01706935

● Tecniche di *machine learning* per la stima della riserva sinistri

[[ML Models](#)]

- ⊕ Wüthrich M.V. (2016). *Machine learning in individual claims reserving*. SSRN Manuscript ID 2867897. Published on *Scandinavian Actuarial Journal* 2018/6, 465-480, 2018. [[← CART](#)]
- ⊕ Wüthrich M.V. (2017). *Neural networks applied to Chain-Ladder reserving*. SSRN Manuscript ID 2966126. To appear in *European Actuarial Journal*. [[← Reti neurali](#)]

⊕ Baudry M., Robert C.Y. (2017). *Non parametric individual claim reserving in insurance*. Workshop on “Data Science in Finance and Insurance”, ISBA (UCL), september 2017. [[← ExtRa-Trees](#)]

⊕ Gabrielli A., Wüthrich M.V. (2018). *An individual claims history simulation machine*. Risks 6/2, 29. [[← Reti neurali](#)]

⊕ Gabrielli A., Richman R., Wüthrich M.V. (2018) *Neural network embedding of the over-dispersed Poisson reserving model*. SSRN Manuscript, ID 3288454. [[← Reti neurali](#)]

● ClaimWatching

⊕ D’Agostino L., De Felice M., Montanucci G., Moriconi F., Salciarini M. (2018). *Machine learning per la riserva sinistri individuale. Un’applicazione R.C. Auto degli alberi di classificazione e regressione*. Alef Technical Report – 18/02 – Roma, Agosto.

← ispirato da e basato su Wüthrich 2016

■ Basi metodologiche

□ Breiman L., Friedman J.H., Olshen R.A., Stone C.J. (1984). *Classification and Regression Trees*. Wadsworth Statistics/Probability Series.

□ Hastie T., Tibshirani R., Friedman J. (2009). *The Elements of Statistical Learning. Data Mining, Inference and Prediction*. 2nd edition. Springer Series in Statistics.

□ Therneau T.M., Atkinson E.J., Foundation M., (2015). *An introduction to recursive partitioning using the RPART routines*. R Vignettes, version of June 29, 2015.

□ Wüthrich M.V., Buser C., (2018). *Data analytics for non-life insurance pricing*. SSRN Manuscript 2870308

Machine learning

Obiettivi

1. Stima della riserva sinistri “di ramo” come somma delle riserve sinistri individuali (dopo integrazione per gli IBNYR)
 - tener conto dell’eterogeneità nelle stime su dati aggregati, incorporare informazione individuale
2. **ClaimWatching**: previsione e controllo del processo di rimborso dei sinistri individuali (tramite il *machine learning*)
 - previsione di eventi a un anno (*early warning*)
 - previsione di eventi multiperiodale
 - previsione di costi fino al run-off (*individual claim reserving*, revisione automatica delle *case reserve*)
 - ...

L’applicazione a rami diversi dall’RCA si ottiene “per semplificazione” (assenza di gestioni multiple)

1.1 ClaimWatching. La previsione e il controllo del processo di rimborso dei sinistri individuali tramite il machine learning

- Finalità:
 1. previsione di eventi a un anno (*early warning*)
 2. previsione di eventi multiperiodale
 3. previsione di costi fino a run-off (*individual claims reserving*, revisione automatica della *case reserve*)
 4. *claims reserving* aggregato come somma di stime di riserva individuali (e integrazione IBNYR)
- Metodo di machine learning: alberi di classificazione e regressione (CART)
- Applicazione: a livello di LoB o subLoB specificata (es.: RCA)
- Dati: informazione storica sul processo di rimborso, e su fatti collegati, a livello di singolo sinistro, con profondità temporale massima possibile e dettaglio massimo possibile su caratteristiche quali-quantitative (es.: accident year i , reporting delay j , tipo gestione CARD/NOCARD, materiale/corporale, non/in-causa, chiuso/aperto, serie pagamenti parziali, livello di case reserve, ...)

1.2 Lo schema generale di “regressione”

Con riferimento a un singolo sinistro, si abbia un insieme di n osservazioni (con “rumore”) indipendenti (i dati):

$$\{(W_1, \mathbf{x}_1), (W_2, \mathbf{x}_2), \dots, (W_n, \mathbf{x}_n)\}$$

dove:

- \mathbf{x} esprime le **feature** del sinistro (covariate, variabili esplicative, *predictors*, variabili indipendenti)
Le \mathbf{x} sono vettori a d dimensioni appartenenti a uno spazio \mathcal{X} (*feature space*)
- W è la **variabile risposta** (*response variable*, variabile dipendente)

Per ogni $h = 1, 2, \dots, n$ vale il modello previsivo:

$$\mathbf{E}[W_h] = \mu(\mathbf{x}_h)$$

dove μ è una generica *funzione di regressione*, che va stimata sui dati

Esempi

Caso elementare: le W_h sono *normali* con media:

$$\mu(\mathbf{x}_h) = \mu_0 + \mu_1 x_{h,1} + \cdots + \mu_d x_{h,d}$$

Claims frequency regression: le W_h sono *Poissoniane* con parametro:

$$\mu(\mathbf{x}_h) = e^{\mu_0 + \mu_1 x_{h,1} + \cdots + \mu_d x_{h,d}}$$

Per es., nell'RCA:

- W : numero di sinistri all'anno (var. discreta)
- x_1 : età del conducente (variabile continua o discreta)
- x_2 : potenza del motore (variabile continua o discreta)
- x_3 : sesso del conducente (variabile categorica)
- x_4 : provincia di residenza (variabile categorica)
- ...

Mean Square Error of Prediction (MSEP)

Data una stima $\hat{\mu}$ di μ , si ottiene la *previsione (prediction)* $\widehat{W} = \widehat{\mathbf{E}}[W] = \hat{\mu}(\mathbf{x})$

Il MSEP è dato da:

$$\mathbf{E} \left[\left(W - \widehat{W} \right)^2 \right] = \underbrace{\left(\mathbf{E}[W] - \mathbf{E}[\widehat{W}] \right)^2}_{\text{Bias}} + \underbrace{\mathbf{Var}(\widehat{W})}_{\text{EstimationVar.}} + \underbrace{\mathbf{Var}(W)}_{\text{ProcessVar.}}$$

→ *bias-variance trade-off*

1.3 Le risposte dei metodi di machine learning

Come procedere se:

- non si ha idea sulla *funzionale* di μ ?
- lo spazio \mathcal{X} ha *dimensionalità* molto elevata?

I metodi di *machine learning* consentono di rispondere in modo flessibile e efficiente

Dato che si è definita una funzione risposta, si tratterà di *supervised learning*

In questa applicazione al loss reserving:

- La metodologia statistica adottata è quella degli *alberi di classificazione/regressione*: classe di modelli non parametrici che “apprendono dai dati la struttura di μ ”
→ *calibrazione automatica sui dati di un modello previsivo non parametrico*
- Il modello previsivo adottato è un *compound model*, ottenuto componendo:
 - un modello previsivo per l'accadimento degli eventi rilevanti (*occurrences*): *frequency model*
 - un modello previsivo per i costi medi (data l'*occurrence*): *severity model*

Complessità dell'R.C. Auto

Per l'RCA andrebbero modellate quattro **gestioni**. Per semplicità si considerano:

- “NoCARD”
- “CARD” := “CARD Gestionaria” – “Forfait Gestionaria” + “CARD Debitrice”

1.3 Previsione di eventi a un anno

Scelta della variabile risposta

Alla data t , e adottando una griglia temporale annua, consideriamo, a livello individuale (quindi per un fissato sinistro), la variabile risposta al tempo $t + 1$

$$W_{t+1} := YNC_{t+1} + 2 \cdot YCA_{t+1} + 4 \cdot Z_{t+1} + 8 \cdot L_{t+1}$$

definita dagli *indicatori di evento* (0/1)

- YNC_{t+1} : “c’è un pagamento NOCARD”
- YCA_{t+1} : “c’è un pagamento CARD”
- Z_{t+1} : “il sinistro è chiuso”
- L_{t+1} : “il sinistro è in causa”

→ Variabile *categorica* unidimensionale

L’obiettivo è ottenere la *distribuzione di probabilità* della variabile aleatoria W_{t+1} (a 16 valori) sulla base di un vettore x_t di covariate (*features*) osservate fino al tempo t

La variabile risposta

Tabella 1 – Struttura della variabile risposta W

variabili					descrizione stato
YNC	YCA	Z	L	W	
0	0	0	0	0	ASP0 – aperto, senza pagamenti e non in causa
1	0	0	0	1	APN0 – aperto, con pagamento NO CARD e non in causa
0	1	0	0	2	APC0 – aperto, con pagamento CARD e non in causa
1	1	0	0	3	APP0 – aperto, con pagamento CARD e NO CARD e non in causa
0	0	1	0	4	CSP0 – chiuso, senza pagamenti e non in causa
1	0	1	0	5	CPN0 – chiuso, con pagamento NO CARD e non in causa
0	1	1	0	6	CPC0 – chiuso, con pagamento CARD e non in causa
1	1	1	0	7	CPP0 – chiuso, con pagamento CARD e NO CARD e non in causa
0	0	0	1	8	ASPL – aperto, senza pagamenti e in causa
1	0	0	1	9	APNL – aperto, con pagamento NO CARD e in causa
0	1	0	1	10	APCL – aperto, con pagamento CARD e in causa
1	1	0	1	11	APPL – aperto, con pagamento CARD e NO CARD e in causa
0	0	1	1	12	CSSL – chiuso, senza pagamento e in causa
1	0	1	1	13	CPNL – chiuso, con pagamento NO CARD e in causa
0	1	1	1	14	CPCL – chiuso, con pagamento CARD e in causa
1	1	1	1	15	CPPL – chiuso, con pagamento CARD e NO CARD e in causa

La struttura della variabile risposta è a 4 componenti binarie, W può assumere 16 possibili valori

Esempi di features

In generale le feature possono essere variabili sia *qualitative* che *quantitative*.

Variabili statiche

- Id – codice identificativo
- i – accident year
- j – reporting delay
- ν – numero d'ordine in (i, j)

Variabili dinamiche

- S_t – cash flow di rimborso pagato in t per il sinistro
- Z_t – indicatore dello stato di “chiusura” (chiuso/aperto) del sinistro in t
- L_t – indicatore dello stato “legale” (in-causa/non-in-causa) del sinistro in t
- $pcum_t$ – pagato cumulato per il sinistro fino all'anno t

In generale, una variabile dinamica può assumere sia il ruolo di variabile esplicativa che quello di variabile risposta

L'indicatore R_t del tipo di case reserve (CARD/NO CARD) e l'ammontare di case reserve $caseR_t$ possono essere considerati, a seconda delle situazioni, sia variabili statiche che dinamiche (cfr. seg.)

L'ipotesi base del modello previsivo

Si definisca il:

time lag ℓ : tempo trascorso tra la data t e l'anno di accadimento del sinistro

Indicando allora con \mathcal{F}_t l'informazione rilevante (σ -algebra) disponibile alla data t , si definisce il modello di regressione:

$$\mathbf{E}[W_{t+1} | \mathcal{F}_t] = \mu_\ell(\mathbf{x}_t)$$

dove la *funzione di regressione* μ dipende dal lag ℓ

→ un modello per ogni lag!

Per dare “struttura” al modello è necessario specificare il ruolo delle variabili temporali rilevanti per il *loss reserving*.

Notazione estesa

Sia:

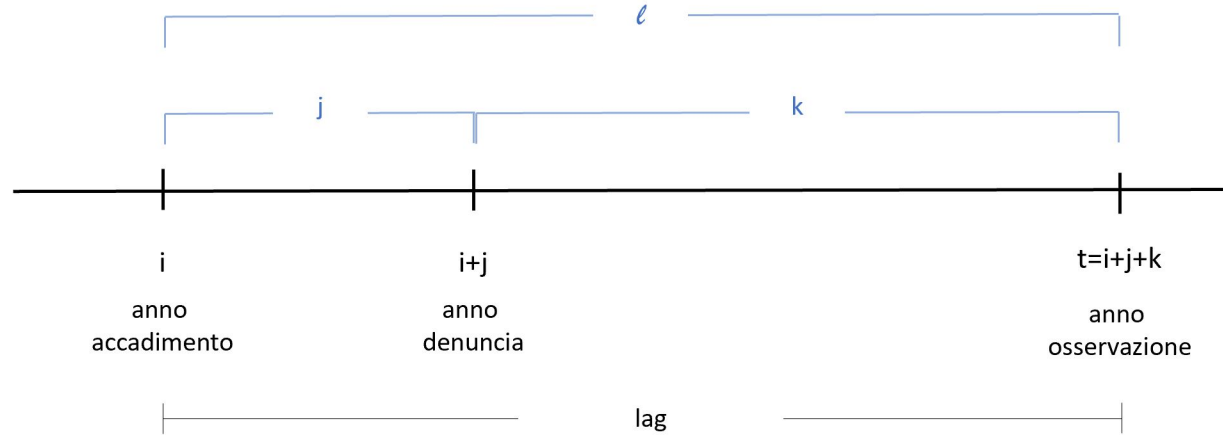
- $i = 1, 2, \dots, I$ – l'indice che denota l'anno di accadimento (*accident year, ay*) dei sinistri osservati fino a t
- $j = 0, 1, \dots, I - i$ – il ritardo di denuncia, in anni (*reporting delay, rd*) (l'anno di denuncia è quindi $i + j$)
- $k = 0, 1, \dots, t - (i + j)$ – il tempo tra la data di osservazione t e la denuncia
- $N_{i,j}$ – il numero di sinistri con ay i e rd j (“blocco di sinistri” (i, j))
- $\nu = 1, 2, \dots, N_{i,j}$ – l'indice del sinistro nel blocco (i, j)

Si ha quindi $t = i + j + k$ e

$$\ell := t - i = j + k = 0, 1, \dots, t - 1$$

(ma per $\ell = t - 1$ la risposta W non è osservata)

Rappresentazione degli indici temporali



Sia allora:

- $\mathcal{C}_{i,j}^{(\nu)}$ – il ν -esimo sinistro del blocco (i, j)
- $v_{i,j|k}^{(\nu)}$ – il corrispondente valore alla data $t = i + j + k$ della generica variabile v (indipendente o dipendente)

Il modello per il lag ℓ si specifica, per ogni i, j, ν , nella forma:

$$\mathbf{E} \left[\underbrace{W_{i,j|k+1}^{(\nu)}}_{t+1} \middle| \underbrace{\mathcal{F}_{i+j+k}}_t \right] = \mu_\ell \left(\underbrace{\mathbf{x}_{i,j|k}^{(\nu)}}_t \right), \quad \ell = 0, 1, \dots, I - 2,$$

con $k = \ell - j$

La struttura “triangolare” dei dati per la stima

Tabella 2 - Struttura dei dati (feature e response)

				Features-Response at lag $\ell = j + k$, $\ell' = \ell + 1$ (step ℓ)			
ld	ay: i	rd: j	ν	$\ell = 0$	$\ell = 1$	$\ell = 2$	$\ell = 3$
1	1	0	1	$(W_{1,0 1}^{(1)}, \mathbf{x}_{1,0 0}^{(1)})$	$(W_{1,0 2}^{(1)}, \mathbf{x}_{1,0 1}^{(1)})$	$(W_{1,0 3}^{(1)}, \mathbf{x}_{1,0 2}^{(1)})$	$(\cdot, \mathbf{x}_{1,0 3}^{(1)})$
2	1	1	1	nyr	$(W_{1,1 1}^{(1)}, \mathbf{x}_{1,1 0}^{(1)})$	$(W_{1,1 2}^{(1)}, \mathbf{x}_{1,1 1}^{(1)})$	$(\cdot, \mathbf{x}_{1,1 2}^{(1)})$
3	1	2	1	nyr	nyr	$(W_{1,2 1}^{(1)}, \mathbf{x}_{1,2 0}^{(1)})$	$(\cdot, \mathbf{x}_{1,2 1}^{(1)})$
4	1	3	1	nyr	nyr	nyr	$(\cdot, \mathbf{x}_{1,3 0}^{(1)})$
5	2	0	1	$(W_{2,0 1}^{(1)}, \mathbf{x}_{2,0 0}^{(1)})$	$(W_{2,0 2}^{(1)}, \mathbf{x}_{2,0 1}^{(1)})$	$(\cdot, \mathbf{x}_{2,0 2}^{(1)})$	\cdot
6	2	1	1	nyr	$(W_{2,1 1}^{(1)}, \mathbf{x}_{2,1 0}^{(1)})$	$(\cdot, \mathbf{x}_{2,1 1}^{(1)})$	\cdot
7	2	2	1	nyr	nyr	$(\cdot, \mathbf{x}_{2,2 0}^{(1)})$	\cdot
8	3	0	1	$(W_{3,0 1}^{(1)}, \mathbf{x}_{3,0 0}^{(1)})$	$(\cdot, \mathbf{x}_{3,0 1}^{(1)})$	\cdot	\cdot
9	3	1	1	nyr	$(\cdot, \mathbf{x}_{3,1 0}^{(1)})$	\cdot	\cdot
10	4	0	1	$(\cdot, \mathbf{x}_{4,0 0}^{(1)})$	\cdot	\cdot	\cdot

Le osservazioni in $t = I = 4$ delle coppie (Features-Response) sono organizzate per lag ($\ell = t - i = j + k$, con j reporting delay e k il tempo che intercorre tra la data di osservazione e la denuncia)

Le risposte dell’“ultima diagonale” non sono state ancora osservate

La stima coi CART

La funzione di regressione μ_ℓ viene stimata in modo non-parametrico con l'approccio degli *alberi di classificazione* (dato che la risposta W è una variabile categorica).

Secondo questo approccio la funzione stimata $\hat{\mu}_\ell$ è *costante a tratti* su una opportuna **partizione** dello spazio \mathcal{X} delle features:

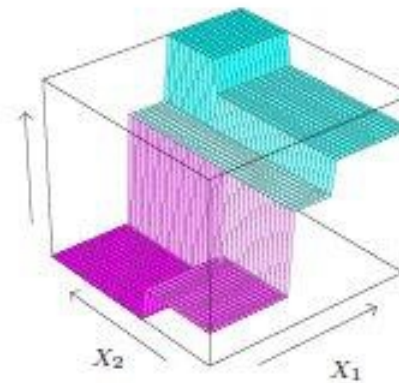
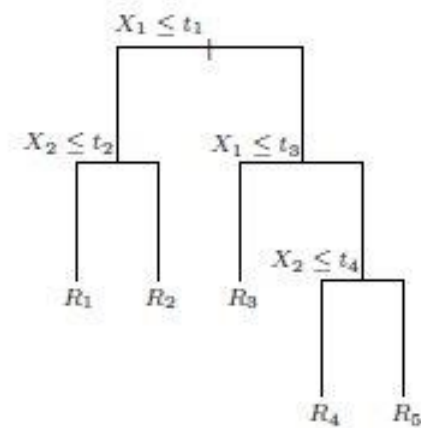
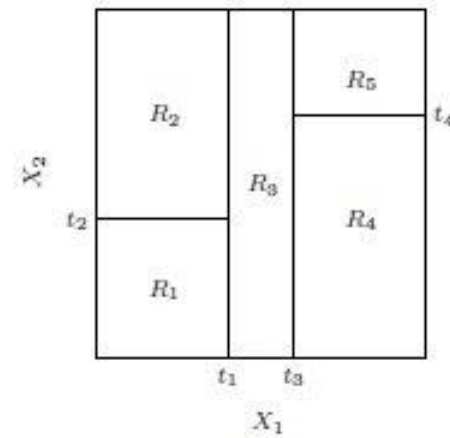
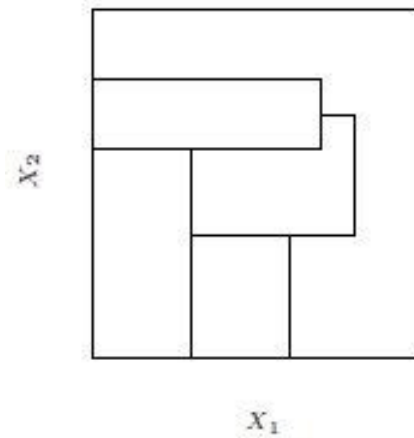
$$\mathcal{P}_\ell := \left\{ \mathcal{R}_\ell^{(1)}, \dots, \mathcal{R}_\ell^{(n_\ell)} \right\}$$

Esistono cioè n_ℓ costanti $\bar{\mu}_\ell^{(r)}$ tali che:

$$\hat{\mu}_\ell \left(\mathbf{x}_{i,j|\ell-j}^{(\nu)} \right) = \sum_{r=1}^{n_\ell} \bar{\mu}_\ell^{(r)} \mathbf{1}_{\left\{ \mathbf{x}_{i,j|\ell-j}^{(\nu)} \in \mathcal{R}_\ell^{(r)} \right\}}$$

Il CART raggruppa nello stesso *iperrettangolo* $\mathcal{R}_\ell^{(r)}$ osservazioni di lag (W, \mathbf{x}) in qualche senso simili tra loro.

Gli iperrettangoli sono le *foglie* (*leaves*) dell'albero, e possono essere interpretate come *classi esplicative*.



Partizione di \mathcal{X} e funzione $\hat{\mu}$ stimata nel caso di un *feature space* bidimensionale (Hastie et al. 2009)

Giustificazione teorica dell'approccio coi CART

Sia $\mathcal{C}_{i,j}^{(\nu)}$ il ν -esimo sinistro del blocco (i, j) e si indichi con $\mathbf{D}_{i,j|k}^{(\nu)}$ il vettore delle *dynamic features* di $\mathcal{C}_{i,j}^{(\nu)}$ osservate fino alla data $t = i + j + k$.

(Hp1) Le v.a. in $(\mathbf{D}_{i,j|k}^{(\nu)})_{i,j,\nu,k}$ di accident year diversi sono indipendenti.

(Hp2) Per ogni accident year i processi $(\mathbf{D}_{i,j|k}^{(\nu)})_{i,j,\nu,k}$ relativi a valori di j e di ν diversi sono indipendenti.
Per la distribuzione di probabilità della risposta a 4 componenti si assume:

$$W_{i,j|k+1}^{(\nu)} | \mathcal{F}_{i+j+k} \sim \text{Categorical} \left(p_{i,j+k}^{(n,c,z,l)}(\mathbf{x}_{i,j|k}^{(\nu)}) \right), \quad (1)$$

dove $p_{i,j+k}^{(n,c,z,l)} : \mathcal{X} \mapsto [0, 1]^{2^4}$ è una funzione di probabilità, e perciò è:

$$\sum_{n,c,z,l \in \{0,1\}} p_{i,j+k}^{(n,c,z,l)}(\mathbf{x}_{i,j|k}^{(\nu)}) = 1.$$

La (1) equivale alla proprietà:

$$\mathbf{P} \left[W_{i,j|k+1}^{(\nu)} = w | \mathcal{F}_{i+j+k} \right] = p_{i,j+k}^{(n,c,z,l)}(\mathbf{x}_{i,j|k}^{(\nu)}), \quad (2)$$

dove l'evento $(W_{i,j|k+1}^{(\nu)} = w)$ con $w \in \{0, \dots, 15\}$ equivale all'evento:

$$\left(\text{YNC}_{i,j|k+1}^{(\nu)} = n, \text{YCA}_{i,j|k+1}^{(\nu)} = c, Z_{i,j|k+1}^{(\nu)} = z, L_{i,j|k+1}^{(\nu)} = l \right) \text{ con } n, c, z, l \in \{0, 1\}.$$

Algoritmo di sviluppo dell'albero

- Ogni *binary split* dello spazio delle feature \mathcal{X} è ottenuto minimizzando una *loss function* scelta come una appropriata *impurity measure* (in questo caso, un indice di Gini)
- In una prima fase viene sviluppato un albero sovrabbondante
- In una seconda fase la lunghezza ottimale dell'albero è ottenuta per “potatura” (*cost-complexity pruning*), con tecniche di *cross validation* (*K-fold validation*)
- Gli iperrettangoli $\mathcal{R}_\ell^{(r)}$ sono le “foglie” del *classification tree*

Esempio di applicazione a dati reali

- Sinistri RCA. Data di osservazione/valutazione: 2015
- Generazioni osservate: dal 2010 al 2015 $\implies i = 1, \dots, I$, con $I = t = 6$
- Osservati solo i sinistri denunciati dal 2013 $\implies j = \max\{4 - i, 0\}, \dots, 6 - i$
(3 “diagonali”)
- Coppie (W, x) osservate per i lag $\ell = 0, \dots, I - 2 = 4$: possibili 5 step di stima [●]
Si costruisce un set di dati per ogni stima: $\mathcal{D}_\ell, \ell = 0, \dots, 4$
- In \mathcal{D}_ℓ l'indice temporale delle variabili dinamiche può andare da 0 a ℓ

Modello frequency per $\ell = 0$:

- si hanno osservazioni di 284.336 sinistri
- si scelgono le features: $j, L0, Z0, YCA0, YNC0, RCA0, RNC0$
(non tutte saranno necessariamente utilizzate)

Albero di classificazione per $\ell = 0$

Si considerano tutti i sinistri osservati con lag $\ell = 0$ (284.336 sinistri):

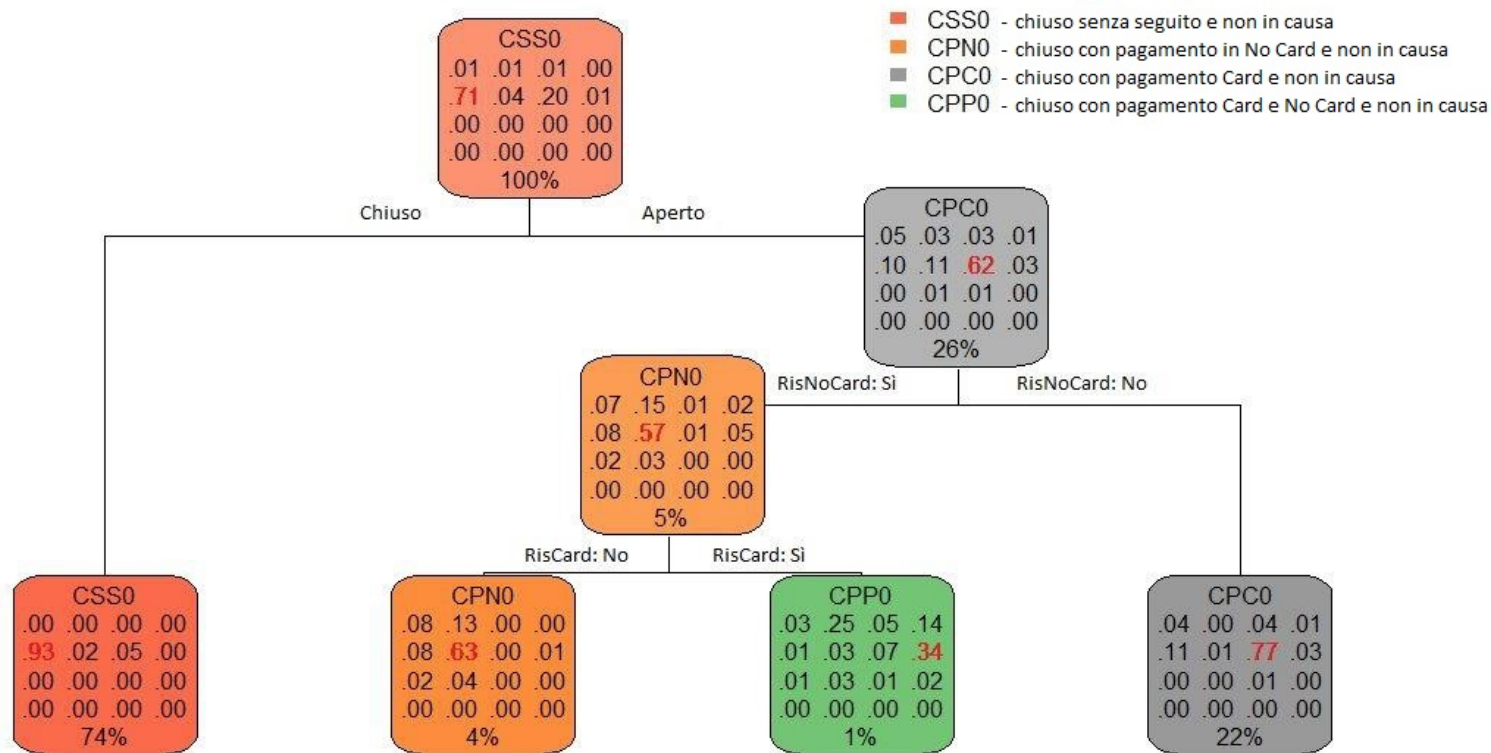


Figura 1 – Albero di classificazione al lag $\ell = 0$.

L'algoritmo indica che la migliore capacità previsiva si ottiene suddividendo i sinistri con lag $\ell = 0$ in 4 classi esplicative. Per ogni classe è stimata la distribuzione di probabilità dei 16 stati

I risultati in forma tabellare

Tabella 3 - Distribuzioni di probabilità stimate dall'algoritmo

nodo	nobs	caratteristiche della foglia	CM	distribuzione di probabilità della risposta							
1	284.336	sinistri omogenei	CSS0	ASP0	APN0	APC0	APP0	CSS0	CPN0	CPC0	CPP0
				0.014	0.0087	0.0084	0.0021	0.71	0.042	0.2	0.0081
				ASPL	APNL	APCL	APP0	CSSL	CPNL	CPCL	CPPL
				0.002	0.0021	0.0017	0.00026	0	0.00019	0.00021	0.00004
2	209.046	sinistri chiusi	CSS0	ASP0	APN0	APC0	APP0	CSS0	CPN0	CPC0	CPP0
				0.024	0.001	0.00076	0.00006	0.93	0.017	0.046	0.00035
				ASPL	APNL	APCL	APP0	CSSL	CPNL	CPCL	CPPL
				0.00094	0.00037	0.00032	0.00001	0	0.00004	0.00004	0
12	12.659	sinistri aperti solo per la partita di danno NoCard	CPN0	ASP0	APN0	APC0	APP0	CSS0	CPN0	CPC0	CPP0
				0.076	0.13	0.0011	0.0044	0.083	0.63	0.0023	0.011
				ASPL	APNL	APCL	APP0	CSSL	CPNL	CPCL	CPPL
				0.017	0.035	0.00024	0.00063	0	0.0035	0	0.00016
13	1.472	sinistri aperti per entrambe le partite di danno	CPP0	ASP0	APN0	APC0	APP0	CSS0	CPN0	CPC0	CPP0
				0.034	0.25	0.052	0.14	0.0054	0.032	0.071	0.34
				ASPL	APNL	APCL	APP0	CSSL	CPNL	CPCL	CPPL
				0.013	0.028	0.0088	0.021	0	0	0	0.0014
7	61.159	sinistri aperti ma non per la partita di danno NoCard	CPC0	ASP0	APN0	APC0	APP0	CSS0	CPN0	CPC0	CPP0
				0.041	0.0035	0.035	0.0052	0.11	0.0052	0.77	0.026
				ASPL	APNL	APCL	APP0	CSSL	CPNL	CPCL	CPPL
				0.0023	0.00036	0.0063	0.00054	0	0.00002	0.00085	0.00013

In rosso sono indicati lo stato modale e la sua probabilità

1.4 Previsione di eventi multiperiodale

Per la previsione degli stati per gli anni successivi al next year si pone un problema multiperiodalità: concatenazione di più modelli uniperiodali

Si possono considerare due approcci alla soluzione del problema:

1) *costruzione di matrici di transizione*

2) *approccio simulativo*

1) L'approccio mediante matrici di transizione è molto oneroso dal punto di vista computazionale

2) Nell'approccio simulativo le risposte prodotte dal modello al lag ℓ vanno ad aggiornare la σ -algebra $\mathcal{F}_{t=i+\ell}$ andando a costituire la σ -algebra $\mathcal{F}_{t+1=i+\ell+1}$, in modo da poter applicare il modello uniperiodale stimato per il lag successivo

1.5 Previsione di costi

Il modello per la severity

La metodologia utilizzata per la previsione di eventi può essere interpretata come un *modello per la frequency*. Per considerare l'ammontare dei costi fino a run-off è necessaria l'aggiunta di un *modello per la severity*

Con riferimento a un fissato sinistro (e usando una notazione semplificata) sia:

- S_t la v.a. costo del rimborso del sinistro effettuato al tempo t
- Y_t l'evento “per il sinistro si ha un pagamento al tempo t ”

Si pone:

$$\mathbf{E}[S_t] = \mathbf{E}[X_t] \mathbf{P}(Y_t = 1)$$

dove X_t è la v.a. condizionata $X_t := S_t | (Y_t = 1)$

La stima della probabilità di pagamento $\mathbf{P}(Y_t = 1)$ è stata già prodotta nella “sezione frequency” È ora richiesta una stima di $\mathbf{E}[X_t]$

La risposta è ora una *variabile continua*: si usa la tecnica CART nella versione dei *regression tree*

Formalmente si ha:

$$\mathbf{E} \left[X_{t+1} \middle| \mathcal{F}_t \right] = \bar{\mu}_\ell^{(r)}(\bar{\mathbf{x}}_t), \quad \bar{\mathbf{x}}_t \in \mathcal{R}_\ell^{(r)}, \quad r = 1, \dots, R_\ell,$$

dove:

- $\mathcal{R}_\ell^{(r)}$ è l'iperrettangolo r -esimo della partizione ottenuta nel modello per la frequency al lag ℓ
- $\bar{\mathbf{x}}_t$ sono le features ritenute significative nel modello per la severity (per es. la *case reserve*)

→

Viene stimato un modello per la severity al lag ℓ per ogni elemento della partizione prodotta dal modello per la frequency al lag ℓ

Albero di regressione per i sinistri con $\ell = 0$ della 2^a “foglia frequency”

Si considerano tutti i sinistri osservati al lag $\ell = 0$ che appartengono al secondo sottoinsieme della partizione della frequency (sinistri aperti, che hanno appostata solo riserva No Card e che hanno avuto pagamento: 10.224 sinistri)

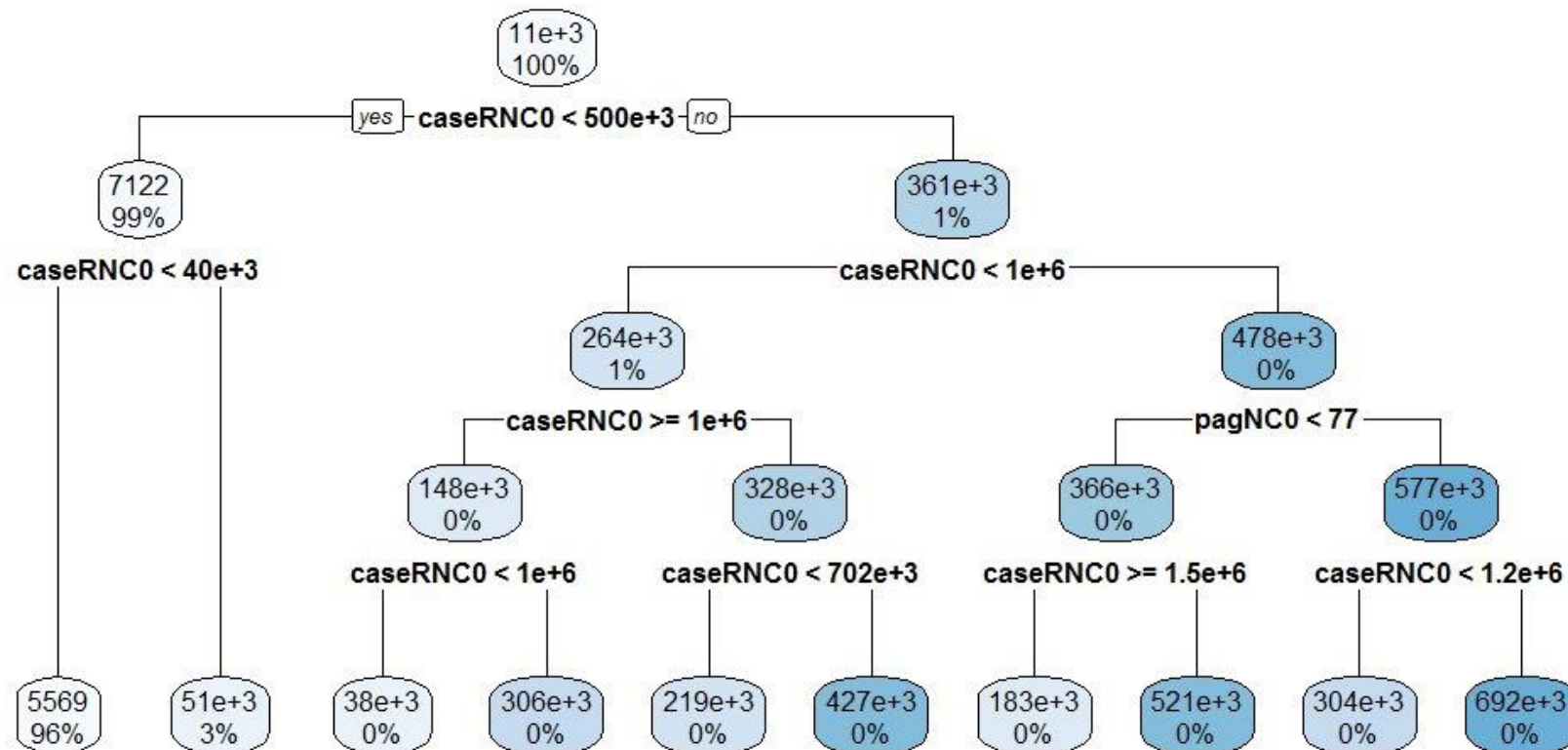
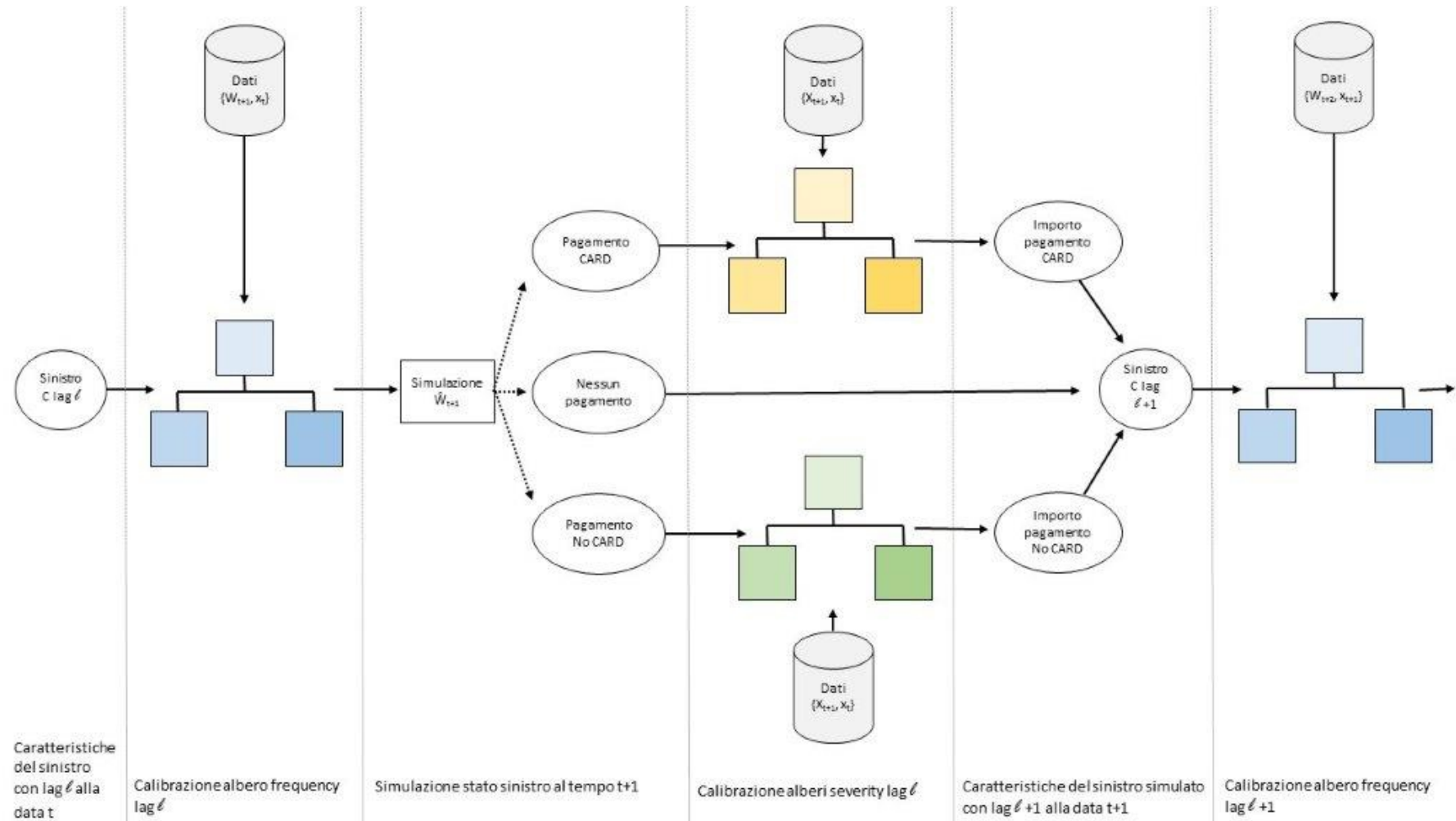


Figura 2 – Albero di regressione al lag $\ell = 0$ per sinistri della seconda classe della frequency.

Lo schema di simulazione multiperiodale



1.6 Previsioni di cost development del singolo sinistro

Si consideri uno specificato sinistro

Features del sinistro:

- $i = I$ – sinistro accaduto nell'anno I
- $j = 0$ – sinistro con reporting delay nullo, cioè l'anno di denuncia coincide con quello di accadimento
- $Z0 = 0$ – sinistro aperto
- $L0 = 0$ – sinistro non in causa
- $YCA0 = 0$ – sinistro senza pagamenti per la partita di danno Card nell'anno I
- $YNC0 = 0$ – sinistro senza pagamenti per la partita di danno No Card nell'anno I
- $RCA0 = 1$ – sinistro per cui è stata appostata riserva Card nell'anno I
- $caseRCA0 = 5.000$ – è stata appostata riserva Card di 5.000 euro nell'anno I
- $RNC0 = 1$ – sinistro per cui è stata appostata riserva No Card nell'anno I
- $caseNC0 = 55.000$ – è stata appostata riserva No Card di 55.000 euro nell'anno I

Traiettorie simulate per la partita di danno Card

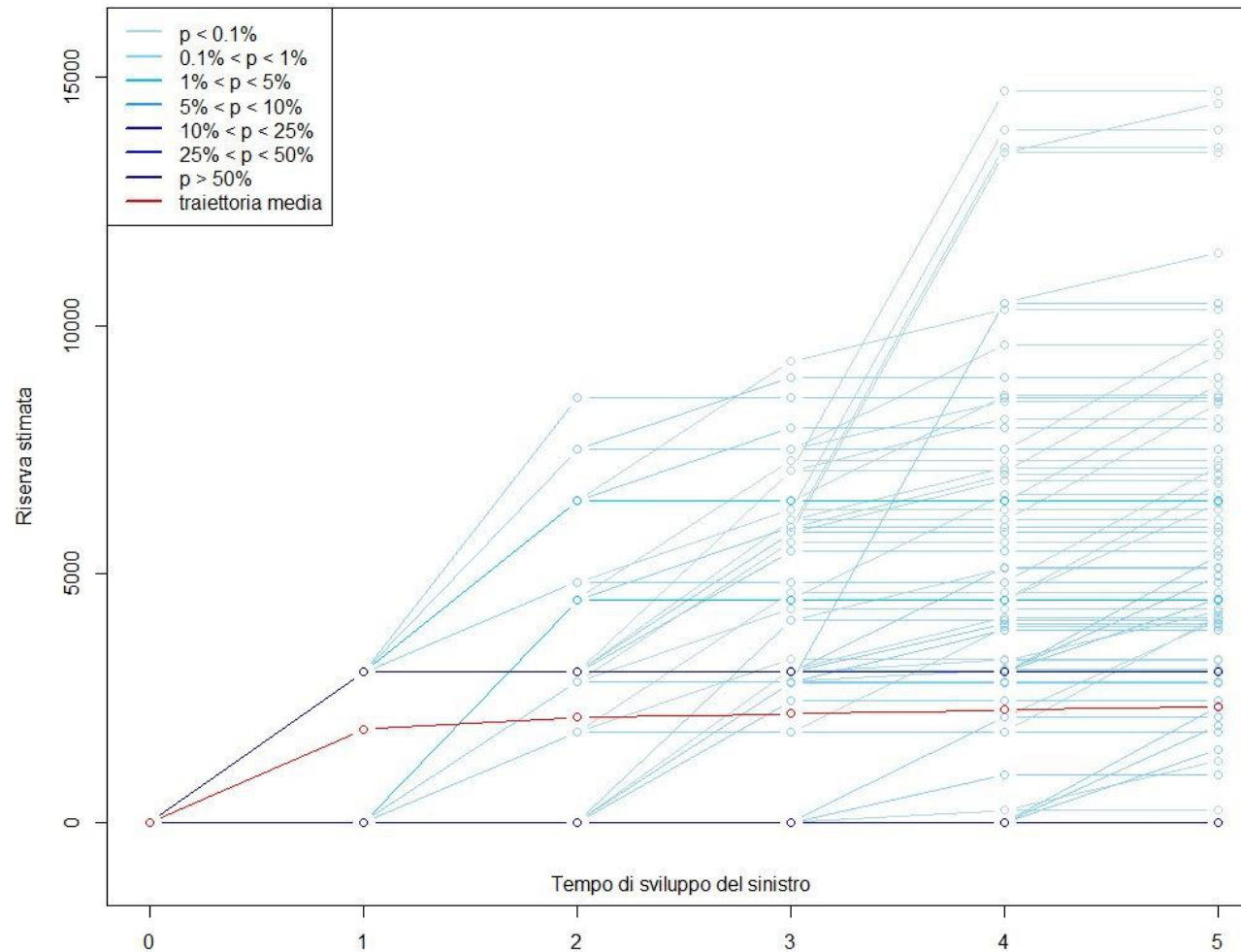


Figura 3 – traiettorie del costo della partita di danno Card, il modello suggerisce di rivedere la riserva Card da 5.000 Euro a 2.322 Euro.

Traiettorie simulate per la partita di danno No Card

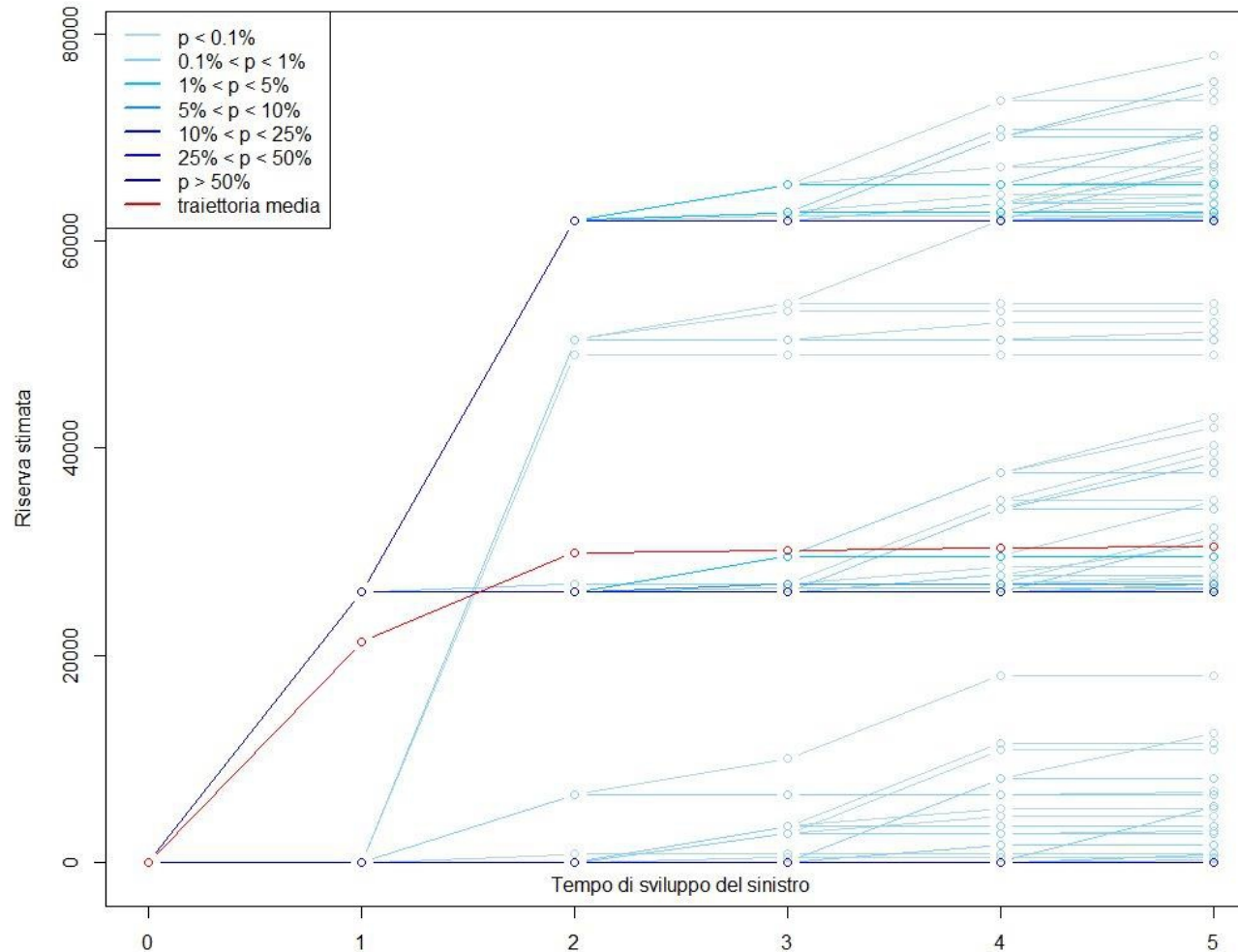


Figura 4 – traiettorie del costo della partita di danno No Card, il modello suggerisce di rivedere la riserva No Card da 55.000 Euro a 30.468 Euro.

Traiettorie simulate per il costo totale del sinistro

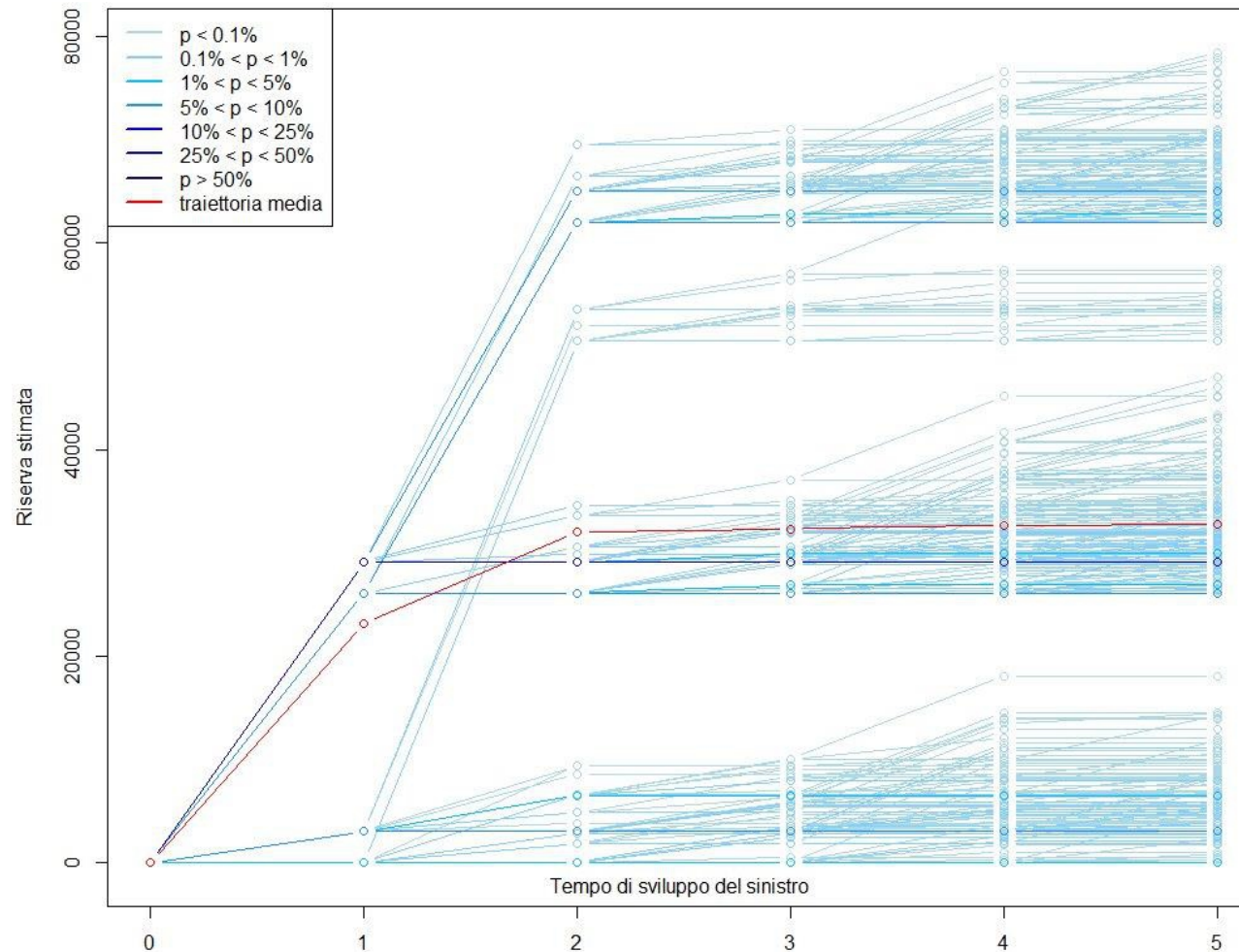


Figura 5 – traiettorie del costo totale del sinistro, il modello suggerisce di rivedere la riserva complessiva da 60.000 Euro a 32.790 Euro.

Interpretazione della case reserve

Tipicamente l'ammontare della case reserve fa parte dell'insieme delle features. Nel modello la case reserve può essere interpretata in due modi:

1. *variabile statica*: se è valorizzata all'istante di denuncia del sinistro e mantenuta costante, a meno di aggiornamenti automatici per i pagamenti parziali
→ funzioni di regressione dipendenti solo dal lag ℓ
applicazione: *stima* della case reserve durante lo sviluppo
2. *variabile dinamica*: se è aggiornata in base a informazioni aggiuntive.

Si distinguono due casi:

- è una variabile non modellizzata
→ funzioni di regressione dipendenti dal lag ℓ e dall'anno di accadimento i
- è una variabile modellizzata
→ funzioni di regressione dipendenti solo dal lag ℓ

applicazione: *revisione* della case reserve durante lo sviluppo

Modellizzare la case reserve dinamica

Si introducono nel modello due “sottosezioni” che si applicano ai sinistri aperti:

- *simil-frequency*. Per determinare il tipo di gestione aperta del sinistro; la variabile risposta è definita come:

$$WR_{t+1} := RCA_{t+1} + 2 \cdot RNC_{t+1},$$

con:

- RCA_{t+1} : “c’è la riserva Card” ,
- RNC_{t+1} : “c’è la riserva No Card” .

- *simil-severity*. Per determinare l’ammontare della riserva, come aspettativa condizionata dato il risultato della sezione *simil-frequency*. Per es., per la riserva Card si pone:

$$\mathbf{E} \left[caseRCA_{t+1} \right] = \mathbf{E} \left[RisCA_{t+1} \right] \mathbf{P} (RCA_{t+1} = 1) ,$$

con:

$$RisCA_{t+1} := caseRCA_{t+1} | (RCA_{t+1} = 1) .$$

1.7 Estensioni e potenziamenti del modello

Riserva RBNS e riserva IBNYR

Il modello fornisce previsioni per i soli sinistri accaduti e denunciati, RBNS *Reported But Not Settled*) e permette per aggregazione di stimare la riserva RBNS complessiva (non scontata) come:

$$R^{RBNS} = \sum_{i=1}^I \sum_{j=0}^{I-i} \sum_{k>I-(i+j)} \sum_{\nu=1}^{N_{i,j}} \mathbf{E} \left[S_{i,j|k}^{(\nu)} \middle| \mathcal{F}_I \right] .$$

Il modello può essere esteso in modo che fornisca una stima del costo dei sinistri IBNYR (*Incurred But Not Yet Reported*), o “IBNR puri”. Con le usuali ipotesi di indipendenza, si può porre:

$$R^{IBNYR} = \sum_{i=1}^I \sum_{j=I-i+1}^J \mathbf{E} [N_{i,j} | \mathcal{F}_I] \mathbf{E} [\bar{S}_{i,j} | \mathcal{F}_I] ,$$

dove:

- J è il valore massimo ipotizzato per rd,
- $\bar{S}_{i,j}$ è il costo medio totale di rimborso stimato per i sinistri del blocco (i, j) .
- L'aspettativa $\mathbf{E} [N_{i,j} | \mathcal{F}_I]$ si può determinare con un chain-ladder applicato ai numeri dei sinistri.
- Per l'aspettativa $\mathbf{E} [\bar{S}_{i,j} | \mathcal{F}_I]$ si può usare l'approccio CART applicato ai sinistri RBNS.

In corso di sperimentazione

- Migliorare le prestazioni del *severity model* utilizzando come *impurity measure* funzioni di verosimiglianza massimizzate rispetto a media e varianza, e utilizzando i parametri stimati nella simulazione.

- Irrobustire il **ClaimWatching** usando metodi “potenziati” di *shallow machine learning*

Criticità dell’approccio CART:

Instabilità: piccole variazioni nei dati possono produrre alberi molto diversi

Lack of smoothness della funzione di regressione ($\hat{\mu}$ è costante a tratti)

—→ metodi potenziati:

- *Bagging*
- *Random Forests*
- *Boosting machines*

(Wüthrich & Bücher 2018)

In God we trust, all others bring data*

*Hastie et al. 2009

Improved Chain Ladder

2.1 Combined Actuarial Neural Net

L'idea: potenziare un modello classico di regressione statistica arricchendolo con informazioni prodotte da una rete neurale opportunamente costruita e “addestrata”

→

immergere/annidare (*embedding/nesting*) il modello classico in una rete neurale così che

- la calibratura della rete sia inizializzata con i parametri stimati del modello classico
- l'output della rete costituisca una miscela (*blending*) di output classico e neurale

In un certo senso, la rete “esplora” i residui del modello classico in modo da produrre un modello con struttura più ricca.

Il metodo viene qui illustrato con riferimento all'**ODP Reserving Model** (gli esempi numerici sono preliminari/illustrativi)

Lo schema può essere applicato al “**pricing by GLM**”

⊕ Gabrielli A., Richman R., Wüthrich M.V. (2018) *Neural network embedding of the over-dispersed Poisson reserving model*. SSRN Manuscript, ID 3288454.

⊕ Wüthrich M.V., Merz M. (2019) *Editorial: Yes, we CANN!* ASTIN Bulletin **49/1**

2.2 Il modello ODP classico

- $Y_{i,j}$: pagati (aggregati) incrementali nell'AY $i = 1, 2, \dots, I$ e nel DY $j = 0, 1, \dots, J$, con $I = J + 1$ ("triangolo")
- Modello "Over-Dispersed Poisson": $Y_{i,j}/\phi \stackrel{ind.}{\sim} Poi(\mu_{i,j}/\phi)$
 ϕ : parametro di iperdispersione
- Struttura *Cross-Classified* (moltiplicativa):

$$\mu_{i,j} = \eta_i \delta_j$$

η_i : exposure parameters

δ_j : development parameters

$$\implies \mathbf{E}[Y_{i,j}] = \mu_{i,j}, \quad \mathbf{Var}(Y_{i,j}) = \phi \mu_{i,j}$$

Modello definito a meno di una costante moltiplicativa:

per $\eta'_i = \eta_i \cdot cost$ e $\delta'_j = \delta_j / cost$ si ottiene $\eta'_i \delta'_j = \eta_i \delta_j = \mu_{i,j}$

- Parametrizzazione ai fini di stima GLM

Canonical link function *logaritmica*. Ponendo $cost = \eta_1 \delta_0$:

$$\log \mu_{i,j} = c + \alpha_i + \beta_j, \quad \text{con } \alpha_1 = \beta_0 = 0.$$

Interpretazione. Se U_i : costi ultimi e π_j : development pattern ($\sum_j \pi_j = 1$), si ha:

$$c = \log (\mathbf{E}[U_1] \pi_0) , \quad \alpha_i = \log \frac{\mathbf{E}[U_i]}{\mathbf{E}[U_1]} , \quad \beta_j = \log \frac{\pi_j}{\pi_0} .$$

- Stima Maximum Likelihood. Minimizzazione della *Poisson deviance statistics*:

$$\mathcal{L}((\mu_{i,j})_{i,j}) = 2 \sum_{i+j \leq I} \left(\mu_{i,j} - Y_{i,j} + Y_{i,j} \log \frac{Y_{i,j}}{\mu_{i,j}} \right)$$

$$\longrightarrow \alpha_i^{ODP}, \beta_j^{ODP}, c^{ODP} \quad (+ \phi^{ODP})$$

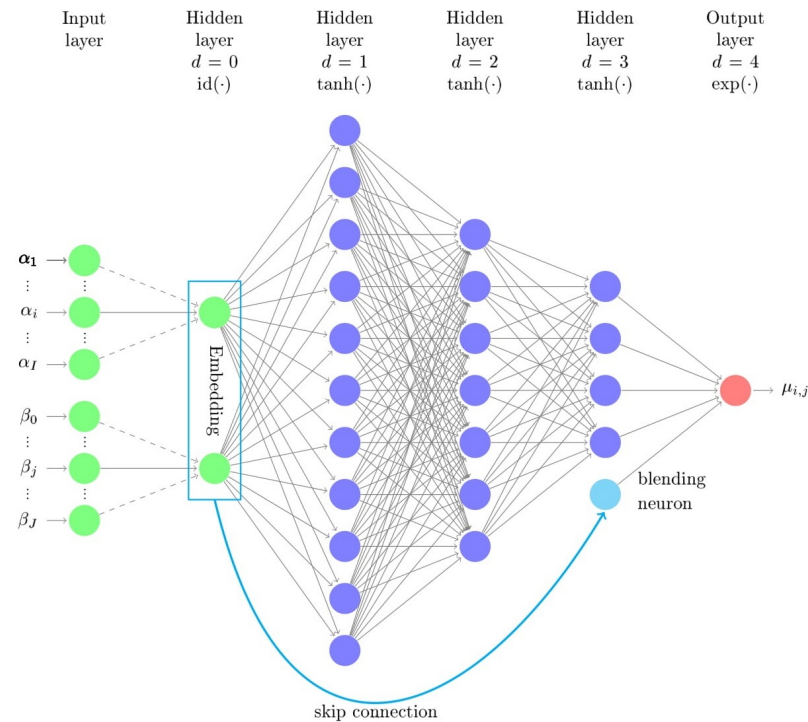
- $Y_{i,j}^{ODP} = \widehat{\mathbf{E}}[Y_{i,j}] = \exp \{ c^{ODP} + \alpha_i^{ODP} + \beta_j^{ODP} \}$

! Su un triangolo le stime di riserva ODP coincidono con le stime chain ladder

“Immersione” dell’ODP in una rete neurale

Struttura tipica della rete feed-forward “ospitante”*

Embedding neural network for [cross-classified] $\mu_{i,j}$ function



→ blended cross classified neural network (bCCNN)

* da Gabrielli et al. 2018

2.3 Struttura algoritmica della rete neurale

- Input

$$\alpha(i) = \alpha_i, \quad \beta(j) = \beta_j$$

- $d = 0$. Embedding layer

Due neuroni $\mathbf{z}^{(0)} = (z_1^{(0)}, z_2^{(0)})'$:

$$\mathbf{z}^{(0)}(i, j) = (\alpha(i), \beta(j))' = (\alpha_i, \beta_j)'$$

- $d = 1, 2, 3$. Layer interni

q_d neuroni $\mathbf{z}^{(d)}$ e funzione di attivazione tangente iperbolica $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$:

$$\mathbf{z}^{(d)}(i, j) = \tanh \left(\mathbf{c}_d + \mathbf{w}'_d \mathbf{z}^{(d-1)} \right)$$

- $d=D=4$. Output

Funzione di attivazione esponenziale + *skip-layer connection*:

$$\mu^{NN}(i, j) = \exp \left\{ \underbrace{\mathbf{c}_4 + \mathbf{w}'_4 \mathbf{z}^3}_{rete} + \underbrace{\alpha_i + \beta_j}_{skip\ conn.} \right\}$$

! Alcuni dettagli architetturali ($D, q_1, \dots, q_D, \dots$) vanno definiti su base empirica

Criterio di calibrazione. Scegliere $(\mathbf{c}_d, \mathbf{w}_d), d = 1, \dots, 4$, in modo da minimizzare la loss function (*deviance loss*)

$$\mathcal{L}((\mu_{i,j})_{i,j}; \phi) = \frac{2}{\phi} \sum_{(i,j) \in \mathcal{T}} \left(\mu_{i,j} - Y_{i,j} + Y_{i,j} \log \frac{Y_{i,j}}{\mu_{i,j}} \right)$$

dove \mathcal{T} è il *training set*.

Metodo. Algoritmo *gradient descent* con retropropagazione (varie versioni)

! Per “annidare” l’ODP classico (già stimato) nella rete neurale:

$$\text{Inizializzazione ODP: } \alpha_i = \alpha_i^{ODP}, \quad \beta_j = \beta_j^{ODP}, \quad c_D = c^{ODP}, \quad \mathbf{w}_D = \mathbf{0}$$

Scelta di efficienza: α_i e β_j *non-trainable*

In che senso la rete “aggiunge struttura”

Come nell'ODP classico, nell'ODP “neurale” i costi ultimi *fitted* si ottengono come:

$$U_i^{NN} = \sum_{j=0}^J \mu^{NN}(i, j)$$

e la stima di riserva è data da:

$$R^{NN} = \sum_{i+j>I} \mu^{NN}(i, j)$$

(naturalmente in generale sarà $U_i^{NN} \neq U_i^{ODP}$).

Tuttavia, mentre nell'ODP classico i development pattern *fitted* hanno la proprietà:

$$\pi_{i,j} := \frac{\mu^{ODP}(i, j)}{\sum_{k=0}^J \mu^{ODP}(i, k)} = \pi_j \text{ per ogni } i$$

nell'ODP neurale si ha:

$$\pi_{i,j} := \frac{\mu^{NN}(i, j)}{\sum_{k=0}^J \mu^N(i, k)} \text{ diverso per ogni } i$$

Addestramento (fitting) della rete e controllo dell'overfitting

I dati vengono suddivisi in un *training set* \mathcal{T} e in un *validation set* \mathcal{V}

- *Fase di training: in-sample fit.* Si effettua su \mathcal{T} una successione di aggiornamenti (“epoche di training”) dell’algoritmo del gradiente a discesa e si controlla la diminuzione della loss function (*training loss*)
- *Fase di validation: out-of-sample analysis.* La rete “addestrata” su \mathcal{T} viene applicata a \mathcal{V} (a produrre la *validation loss*) e si sceglie il numero ottimale di epoche come trade-off tra capacità predittiva e overfitting

! I parametri degli algoritmi di calcolo vanno opportunamente regolati (learning rate, batch size, momentum coefficient, dropout rate, ...)

2.4 Esempio (preliminare)

Triangolo complessivo di tutti i sinistri R.C. Generale dal 2001 al 2017

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2001	8.768.784	13.923.004	8.834.617	6.156.302	5.568.257	5.182.596	5.057.967	5.949.687	3.688.419	3.300.806	2.114.233	3.285.720	1.555.429	2.150.057	2.677.618	1.056.846	742.358
2002	10.238.842	17.443.012	8.268.280	7.080.799	5.987.060	5.211.553	4.025.699	3.230.473	3.174.280	3.399.462	1.810.427	1.111.049	1.162.794	1.855.831	1.020.384	202.286	.
2003	11.611.189	20.396.289	11.446.708	7.626.567	4.804.765	4.840.280	5.923.252	5.418.688	6.449.949	3.909.807	3.268.034	2.118.776	1.617.815	3.567.698	1.477.697	.	.
2004	14.672.040	22.869.165	14.369.665	8.610.455	7.475.484	5.015.438	4.324.296	3.644.509	5.269.134	2.274.292	2.212.727	2.688.437	1.540.310	1.936.406	.	.	.
2005	13.075.735	22.906.605	13.499.339	11.473.692	6.632.824	5.422.158	4.696.014	5.382.841	2.546.516	3.780.981	2.320.898	3.439.403	761.408
2006	14.742.576	26.168.376	14.130.965	8.376.871	7.004.426	6.889.971	3.553.458	2.855.358	6.168.221	2.459.501	1.836.757	3.515.301
2007	15.830.319	27.991.287	12.866.490	6.844.164	5.731.777	6.182.518	6.559.940	3.978.143	3.678.291	1.845.773	4.691.798
2008	18.689.936	30.554.387	15.678.526	11.612.059	9.930.968	7.585.253	6.366.160	3.739.590	4.839.996	2.088.156
2009	18.830.364	25.383.904	14.427.998	10.561.956	6.240.546	7.834.186	9.061.553	4.094.478	2.789.304
2010	16.947.788	25.271.343	15.016.364	8.144.696	9.331.897	7.306.256	6.640.845	3.653.187
2011	16.497.707	27.794.385	14.847.821	8.766.796	8.547.907	4.158.162	4.023.452
2012	18.189.920	29.525.946	14.451.139	8.849.410	8.751.208	5.078.704
2013	16.639.954	28.378.449	13.354.095	7.384.413	6.696.535
2014	16.503.783	25.651.179	14.154.559	11.704.700
2015	16.393.891	23.360.512	10.756.246
2016	16.680.801	26.066.839
2017	16.330.016

Il training set \mathcal{T} e il validation set \mathcal{V} sono stati costruiti ordinando per anno di accadimento la totalità dei sinistri individuali e includendo quelli di posto dispari (pari) nel training (validation) set

Training set \mathcal{T} dal 2001 al 2017

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2001	4.436.003	6.524.210	4.286.367	2.874.892	2.998.116	3.229.559	2.862.270	3.631.067	1.023.666	1.273.606	899.733	1.661.117	870.522	448.365	784.832	928.588	285.857
2002	5.282.691	8.834.613	4.330.933	2.981.284	3.936.370	3.076.906	2.341.470	1.378.639	1.948.824	2.028.933	560.633	367.763	629.804	1.109.612	96.596	116.765	
2003	5.652.279	10.457.844	4.902.556	3.382.493	2.413.344	2.317.923	2.742.075	2.579.019	3.284.219	2.528.493	1.155.531	1.501.610	1.211.520	1.292.428	510.333		
2004	7.057.132	11.342.871	6.147.019	4.458.542	4.632.859	2.669.369	2.519.787	1.797.562	1.966.628	1.346.780	1.258.960	1.660.057	600.330	1.680.107			
2005	6.614.345	11.075.399	6.824.167	6.182.467	3.250.144	2.875.484	2.468.659	3.342.268	1.319.958	2.787.741	1.452.326	2.183.641	187.446				
2006	7.284.147	12.524.236	6.694.369	3.553.516	3.497.322	4.272.624	1.742.211	1.130.926	2.523.153	1.681.019	859.824	1.881.439					
2007	7.910.968	14.287.755	6.555.134	2.904.514	2.416.638	3.256.090	2.778.561	2.554.501	2.071.647	795.747	1.701.158						
2008	9.228.163	16.436.512	8.351.486	6.213.489	6.719.184	4.786.026	3.842.625	1.406.761	2.054.169	1.084.618							
2009	9.561.012	12.616.095	7.375.124	5.732.968	2.930.796	3.546.326	4.067.406	2.897.081	1.412.753								
2010	8.359.384	12.716.880	7.441.109	3.940.827	5.153.912	3.581.127	3.153.424	1.310.732									
2011	7.994.453	12.518.854	7.658.623	4.377.390	4.171.405	1.704.739	1.222.823										
2012	8.999.948	14.619.359	6.384.423	3.756.619	3.671.505	2.877.027											
2013	8.181.130	15.655.467	6.099.860	4.262.324	2.607.969												
2014	8.413.399	11.785.325	7.024.383	4.764.038													
2015	8.145.241	11.475.846	4.281.630														
2016	7.434.132	12.687.504															
2017	7.934.642																

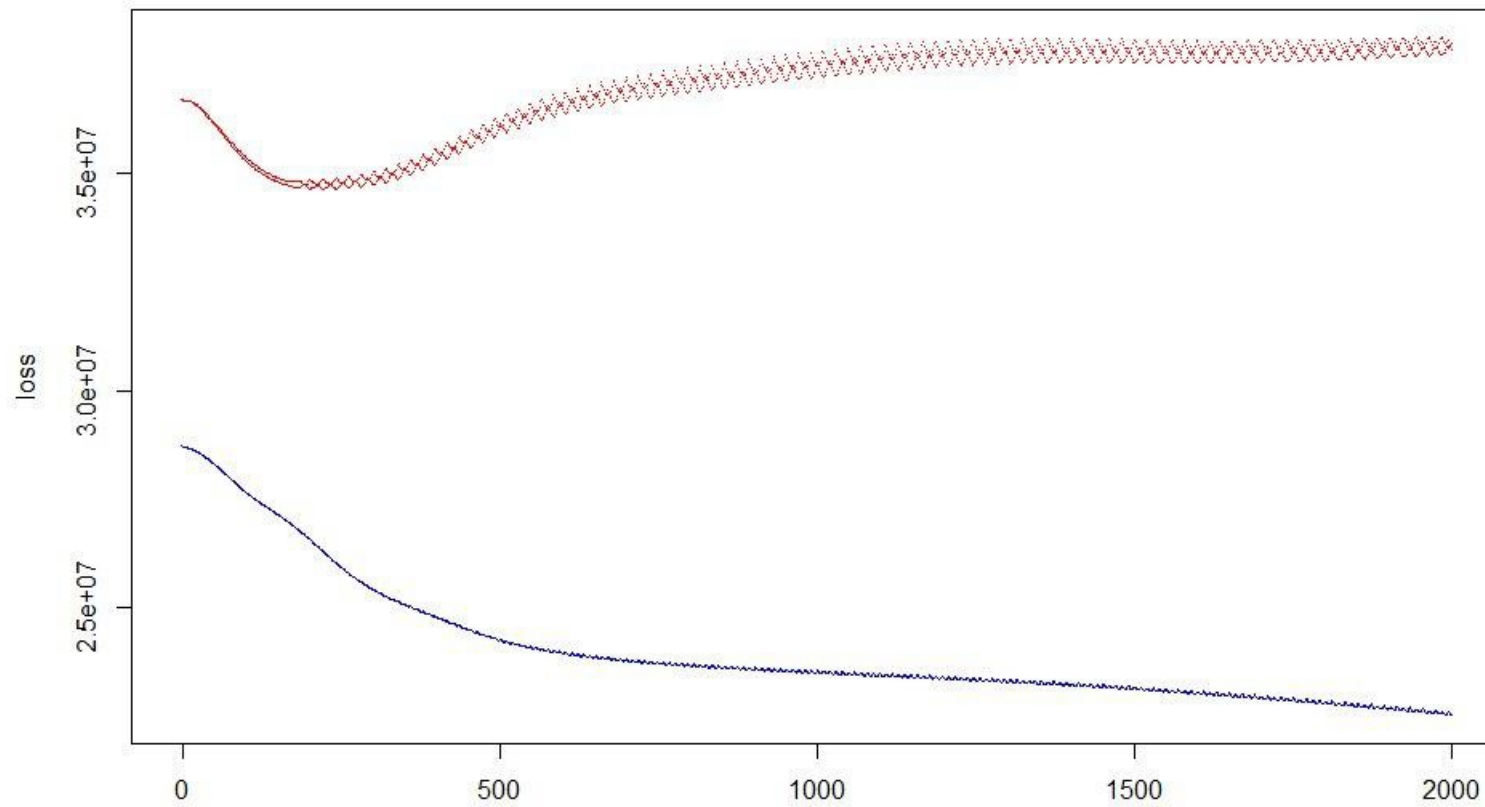
Validation set \mathcal{V} dal 2001 al 2017

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2001	4.332.781	7.398.794	4.548.250	3.281.410	2.570.141	1.953.036	2.195.696	2.318.621	2.664.753	2.027.200	1.214.500	1.624.603	684.907	1.701.692	1.892.786	128.259	456.501
2002	4.956.151	8.608.399	3.937.347	4.099.515	2.050.690	2.134.647	1.684.230	1.851.835	1.225.456	1.370.530	1.249.794	743.287	532.990	746.219	923.787	85.521	
2003	5.958.910	9.938.444	6.544.153	4.244.073	2.391.421	2.522.357	3.181.177	2.839.670	3.165.730	1.381.314	2.112.503	617.166	406.295	2.275.271	967.364		
2004	7.614.908	11.526.294	8.222.646	4.151.913	2.842.625	2.346.069	1.804.509	1.846.947	3.302.506	927.512	953.767	1.028.380	939.980	256.300			
2005	6.461.390	11.831.206	6.675.173	5.291.224	3.382.681	2.546.674	2.227.355	2.040.573	1.226.559	993.240	868.572	1.255.763	573.963				
2006	7.458.428	13.644.141	7.436.596	4.823.355	3.507.104	2.617.347	1.811.247	1.724.432	3.645.069	778.482	976.933	1.633.862					
2007	7.919.350	13.703.532	6.311.356	3.939.650	3.315.139	2.926.427	3.781.379	1.423.642	1.606.644	1.050.027	2.990.639						
2008	9.461.773	14.117.876	7.327.040	5.398.570	3.211.784	2.799.227	2.523.535	2.332.829	2.785.827	1.003.539							
2009	9.269.353	12.767.809	7.052.873	4.828.987	3.309.750	4.287.860	4.994.147	1.197.396	1.376.550								
2010	8.588.404	12.554.463	7.575.256	4.203.870	4.177.985	3.725.129	3.487.421	2.342.455									
2011	8.503.253	15.275.531	7.189.198	4.389.405	4.376.502	2.453.424	2.800.629										
2012	9.189.972	14.906.587	8.066.716	5.092.791	5.079.703	2.201.676											
2013	8.458.824	12.722.981	7.254.235	3.122.088	4.088.566												
2014	8.090.384	13.865.854	7.130.176	6.940.662													
2015	8.248.650	11.884.666	6.474.616														
2016	9.246.670	13.379.335															
2017	8.395.374																

Parametri ODP sul training set

i	α_i^{ODP}	j	β_j^{ODP}
1	0	0	0
2	0.00744	1	0.48373
3	0.18401	2	-0.17927
4	0.26290	3	-0.57911
5	0.31846	4	-0.70558
6	0.27530	5	-0.85552
7	0.30278	6	-1.01039
8	0.57085	7	-1.21457
9	0.42860	8	-1.32231
10	0.38147	9	-1.44723
11	0.29423	10	-1.79851
12	0.38250	11	-1.46638
13	0.38163	12	-2.23699
14	0.35719	13	-1.71270
15	0.21604	14	-2.55277
16	0.32055	15	-2.36984
17	0.35397	16	-2.96948
intercetta:	$c^{ODP} =$	15.53277	

Andamento di training loss e validation loss

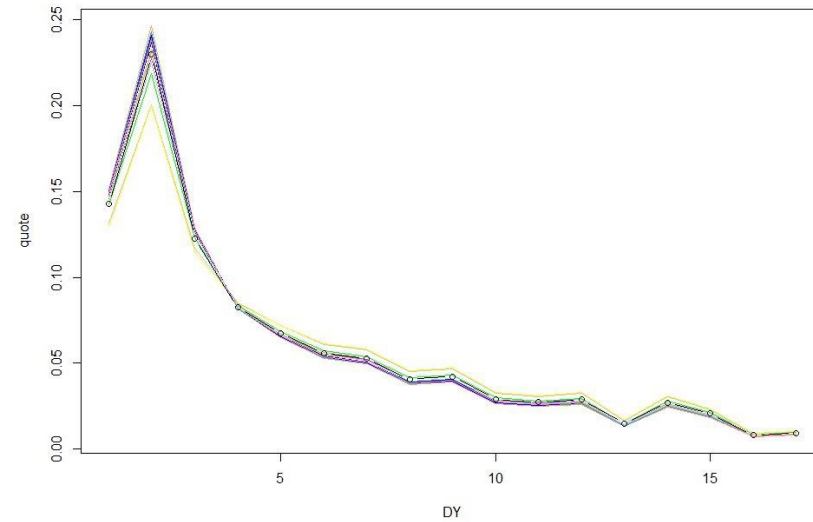
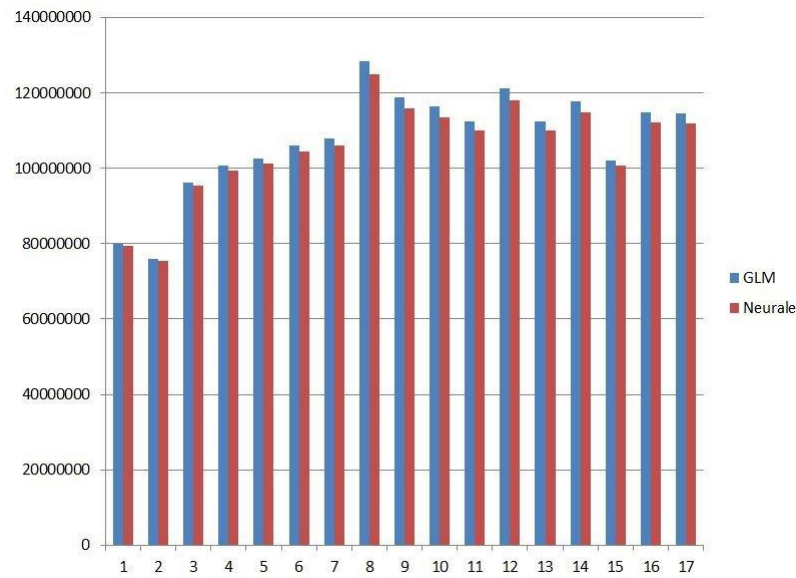


Parametri ODP sul triangolo completo e modelli a confronto

i	α_i^{ODP}	j	β_j^{ODP}	f_j
1	0	0	0	1
2	-0.05241	1	0.47710	2.61140
3	0.18368	2	-0.15046	1.32947
4	0.23015	3	-0.54939	1.16628
5	0.24856	4	-0.75174	1.11646
6	0.28248	5	-0.93757	1.08662
7	0.29875	6	-0.99735	1.07509
8	0.47367	7	-1.25586	1.05393
9	0.39476	8	-1.21941	1.05307
10	0.37427	9	-1.59601	1.03458
11	0.33999	10	-1.65402	1.03154
12	0.41487	11	-1.60028	1.03227
13	0.34005	12	-2.28059	1.01583
14	0.38698	13	-1.66557	1.02883
15	0.24326	14	-1.93792	1.02134
16	0.36118	15	-2.87105	1.00822
17	0.35876	16	-2.73243	1.00936
intercetta:	$c^{ODP} =$	16.24977		

ODP	Classico	“Neurale”
Pagato previsto	469.551.824	443.522.596
Deviance Loss	33.857.834	30.592.134

Confronto tra valori *fitted* dei modelli: costi ultimi U_i e development pattern $\pi_{i,j}$



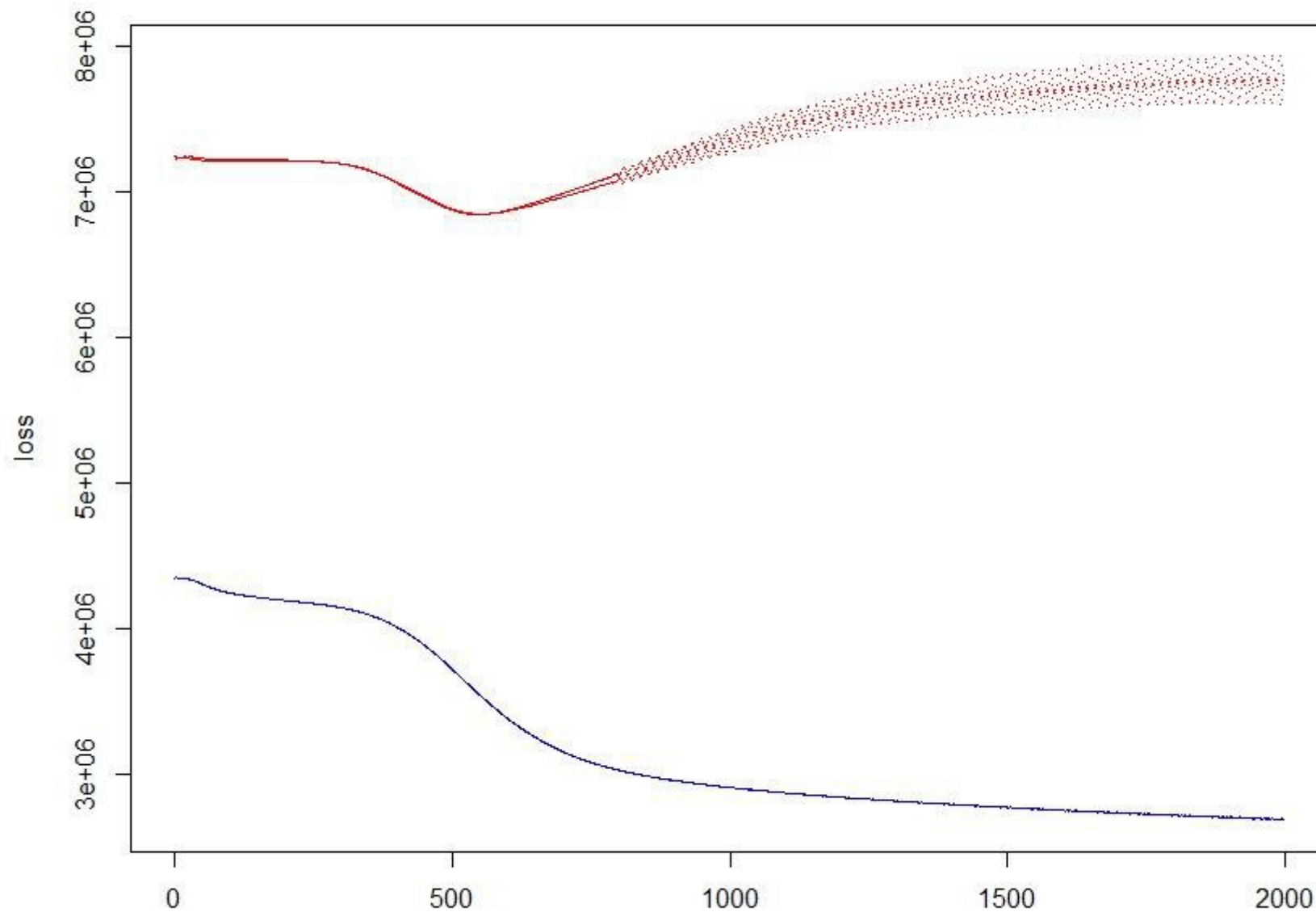
Esempio con backtesting (preliminare)

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2001	8.768.784	13.923.004	8.834.617	6.156.302	5.568.257	5.182.596	5.057.967	5.949.687	3.688.419	3.300.806	2.114.233	3.285.720	1.555.429	2.150.057	2.677.618	1.056.846	742.358
2002	10.238.842	17.443.012	8.268.280	7.080.799	5.987.060	5.211.553	4.025.699	3.230.473	3.174.280	3.399.462	1.810.427	1.111.049	1.162.794	1.855.831	1.020.384	202.286	.
2003	11.611.189	20.396.289	11.446.708	7.626.567	4.804.765	4.840.280	5.923.252	5.418.688	6.449.949	3.909.807	3.268.034	2.118.776	1.617.815	3.567.698	1.477.697	.	.
2004	14.672.040	22.869.165	14.369.665	8.610.455	7.475.484	5.015.438	4.324.296	3.644.509	5.269.134	2.274.292	2.212.727	2.688.437	1.540.310	1.936.406	.	.	.
2005	13.075.735	22.906.605	13.499.339	11.473.692	6.632.824	5.422.158	4.696.014	5.382.841	2.546.516	3.780.981	2.320.898	3.439.403	761.408
2006	14.742.576	26.168.376	14.130.965	8.376.871	7.004.426	6.889.971	3.553.458	2.855.358	6.168.221	2.459.501	1.836.757	3.515.301
2007	15.830.319	27.991.287	12.866.490	6.844.164	5.731.777	6.182.518	6.559.940	3.978.143	3.678.291	1.845.773	4.691.798
2008	18.689.936	30.554.387	15.678.526	11.612.059	9.930.968	7.585.253	6.366.160	3.739.590	4.839.996	2.088.156
2009	18.830.364	25.383.904	14.427.998	10.561.956	6.240.546	7.834.186	9.061.553	4.094.478	2.789.304
2010	16.947.788	25.271.343	15.016.364	8.144.696	9.331.897	7.306.256	6.640.845	3.653.187
2011	16.497.707	27.794.385	14.847.821	8.766.796	8.547.907	4.158.162	4.023.452
2012	18.189.920	29.525.946	14.451.139	8.849.410	8.751.208	5.078.704
2013	16.639.954	28.378.449	13.354.095	7.384.413	6.696.535
2014	16.503.783	25.651.179	14.154.559	11.704.700
2015	16.393.891	23.360.512	10.756.246
2016	16.680.801	26.066.839
2017	16.330.016

Si ripete il calcolo considerando gli anni dal 2001 al 2009 con l'obiettivo di confrontare la riserva stimata dai modelli con il pagato effettivamente osservato (triangolo inferiore).

Il training set e il validation set sono ottenuti in maniera analoga sui rispettivi triangoli 17×17

Andamento di training loss e validation loss



Parametri ODP sul triangolo completo dal 2001 al 2009 e i modelli a confronto

i	α_i^{ODP}	j	β_j^{ODP}
1	0	0	0
2	0.00338	1	0.52668
3	0.19341	2	-0.06869
4	0.37247	3	-0.40181
5	0.38483	4	-0.66008
6	0.43475	5	-0.79656
7	0.49187	6	-0.72680
8	0.64871	7	-0.75057
9	0.67817	8	-0.95218
intercetta:		$c^{ODP} =$	16.07282

Pagato effettivo	245.921.129	
ODP	Classico	“Neurale”
Pagato previsto	315.251.572	284.580.804
Deviance loss	4.290.391	3.662.883

Modelli a confronto per la LoB MTL (dal 2007 al 2012):

Pagato effettivo	657.108.526	
ODP	Classico	“Neurale”
Pagato previsto	741.060.546	730.096.374
Deviance loss	32.967.857	12.248.984

Modelli a confronto per la LoB FDP (dal 2001 al 2009):

Pagato effettivo	118.022.732	
ODP	Classico	“Neurale”
Pagato previsto	153.201.309	149.912.682
Deviance loss	10.494.763	3.672.077

Conclusioni

1. Se l' "immersione" del modello comporta un sostanziale miglioramento della *deviance statistics* dovrebbe essere usato, altrimenti si dovrebbe adottare il modello tradizionale
2. Se il modello tradizionale presenta un buon *fitting*, ci si attende una veloce convergenza del *gradient descent algorithm* poiché è usato come *starting point*
3. È possibile rendere i risultati del modello tradizionale *trainable*, interpretando il coefficiente come *credibility weight* assegnato al modello tradizionale
4. L' "immersione" di modelli in una rete neurale può essere applicato a una ampia varietà di modelli attuariali
5. È possibile applicare la rete neurale per "immergere" più modelli in differenti *skip connection*, facendoli "competere" tra loro