



Il centro Interdipartimentale di Ricerca di Statistica Robusta <http://rosa.unipr.it> e



Il Dipartimento di Scienze Economiche e Aziendali <http://sea.unipr.it> nell'ambito del corso di laurea in Economia e Management (curriculum in Economia Digitale) e del corso di laurea magistrale in Finanza e Risk Management sono lieti di annunciare il seguente seminario:

Data cleaning, data preparation ed individuazione degli outlier per applicazioni finanziarie e di antifrode

MER 28 novembre 2018, ore 11.00-13.00

Aula I del Dipartimento di Scienze Economiche e Aziendali

Relatori

Francesca Perino Mathworks Italia	Emmanuele Sordini Joint Research Centre della Commissione Europea
--------------------------------------	---

ABSTRACT

Preparazione dei dati

Un buon processo di preparazione dei dati rappresenta una fase cruciale per la costruzione di un modello predittivo e si focalizza sull'organizzare e preparare i dati al fine di ottenere il massimo beneficio dall'analisi di tali dati. Ciò permette di far risparmiare tempo nella ricerca di informazioni in quanto i dati saranno più consistenti, validi e non presenteranno errori di nessun genere. In questa sessione vedremo come in MATLAB sia possibile automatizzare il processo di data cleaning e data preparation, sia su dati numerici che testuali.

Esempi utilizzati:

- Analisi fondamentale su titoli S&P 100
- Sentiment Analysis di un'informativa SEC (SEC filings)
- Calcolo di indicatori tecnici su una base molto grande di dati intraday

Individuazione degli outliers multivariati

Ai giorni nostri è universalmente accettato, e non solo tra gli statistici, che la presenza di pochi dati inconsistenti con i rimanenti può alterare radicalmente il risultato di un'analisi statistica e le conseguenze che da essa se ne traggono. In fondo, anche il famoso pollo di Trilussa non fa altro che mettere in evidenza l'incapacità della media aritmetica – tipico stimatore non “robusto” – di sintetizzare adeguatamente il risultato di una rilevazione in cui alcuni individui agiscono in modo molto differente dagli altri. Il tentativo di proteggersi da osservazioni “non rappresentative” del comportamento generale, solitamente dette anomale (o outlier), ha portato gli statistici a proporre una moltitudine di metodi per la loro identificazione. Gli outlier non sono necessariamente osservazioni da buttare. Un campo di analisi in cui gli outlier sono al centro dell'attenzione è quello dell'identificazione di frodi. Una quota importante delle frodi commesse annualmente nell'Unione Europea è perpetrata ai danni degli interessi finanziari degli Stati Membri e dell'Unione stessa da aziende operanti nel commercio intracomunitario ed internazionale. Il seminario si propone (i) di introdurre, senza pretese di completezza matematica, un problema di regressione e un paio di possibili soluzioni robuste, (ii) di estendere il concetto al rilevamento di frodi in dati del commercio internazionale e, infine, (iii) di illustrare come questi metodi sono concretamente utilizzati da operatori dell'antifrode sotto forma di un servizio web. Le illustrazioni faranno uso di strumenti software sviluppati nel progetto FSDA (*Flexible Statistics for Data Analysis*), frutto di una collaborazione proficua tra L'università di Parma e il Centro Comune di Ricerca della Commissione Europea, situato a Ispra (Varese).

Per ulteriori informazioni si prega di contattare

Marco Riani (mriani@unipr.it) oppure Simona Sanfelici (simona.sanfelici@unipr.it)