



WORKSHOP

MACHINE LEARNING IN STATA: UN'INTRODUZIONE

Roma, 21-22 Settembre 2020

Gli ultimi anni hanno visto una disponibilità senza precedenti di informazioni su fenomeni sociali, economici e relativi alla salute. Ricercatori, professionisti e responsabili delle politiche hanno oggi accesso a enormi database (i cosiddetti *Big Data*) su persone, aziende e istituzioni, dispositivi cellulari, web, satelliti, ecc., con sempre maggiore dettaglio.

Il *machine learning* (o “apprendimento automatico”) è un approccio relativamente nuovo all’analisi dei dati, che si colloca nell’intersezione tra statistica, informatica e intelligenza artificiale. Il suo obiettivo principale è quello di trasformare le informazioni in conoscenza e valore, “lasciando che i dati parlino da soli”.

A tal fine, il *machine learning* limita le ipotesi preliminari sulla struttura dei dati e fa affidamento su una filosofia che supporta lo sviluppo di algoritmi, di procedure computazionali e d’ispezione grafica dei risultati più che su assunzioni analitiche e soluzioni algebriche. Computazionalmente irrealizzabile fino a pochi anni fa, il *machine learning* è un prodotto dell’era del computer, della potenza di calcolo e della capacità di apprendimento delle macchine odierne, dello sviluppo dell’hardware e del continuo sviluppo del software.

Questo corso è una introduzione alle tecniche di *machine learning* che utilizzano Stata. Stata possiede oggi vari pacchetti per eseguire il *machine learning* che sono tuttavia poco conosciuti da molti suoi utenti. L’obiettivo, è pertanto colmare questa lacuna rendendo i partecipanti più familiari (e meglio informati) sul potenziale di Stata per trarre conoscenza e valore dai dati, possibilmente di grandi dimensioni e “rumorosi”. L’approccio all’insegnamento sarà principalmente basato sul linguaggio grafico e sull’intuizione più che sull’algebra. Le lezioni si avvarranno di esempi sia simulati che reali, e permetteranno di bilanciare equamente sessioni teoriche e sessioni pratiche. In linea con la nostra filosofia di formazione, le sessioni teoriche sono affiancate da illustrazioni pratiche ed esempi provenienti da diverse discipline sia sociali che biomediche, in cui il docente chiarisce le limitazioni e i punti di forza di ogni metodologia, nonché i criteri per la scelta e l’implementazione dello strumento di analisi statistica più appropriato per il problema oggetto di studio.

Al termine, i partecipanti avranno una migliore comprensione del potenziale di Stata per eseguire il *machine learning*, diventando così in grado di padroneggiare compiti di ricerca che includono, tra gli altri: (i) rilevamento d’importanza dei fattori, (ii) estrazione segnale-rumore, (iii) corretta specificazione del modello, (iv) classificazione, sia da un punto di vista di *data mining* che di approccio causale.

CODICE WORKSHOP

I-EF35

DATA E LUOGO

Roma, 21-22 Settembre 2020

DESTINATARI

Il corso è di interesse per ricercatori e analisti in economia, medicina, marketing e scienze sociali che desiderano acquisire gli strumenti fondamentali per implementare l’approccio di *machine learning* sui così detti *Big Data*.

MACHINE LEARNING IN STATA: UN'INTRODUZIONE

REQUISITI RICHIESTI

Conoscenza di base di econometria/statistica e del Software Stata.

PROGRAMMA

SESSIONE I: LE BASI DEL MACHINE LEARNING

1. *Machine Learning*: definizione, logica, utilità
2. Apprendimento supervisionato vs. non supervisionato
3. Problemi di regressione e di classificazione
4. Inferenza e previsione
5. Errore di campionamento ed errore di specificazione
6. La fondamentale non-identificabilità di $E(y|x)$
 - Modelli parametrici e non parametrici
 - Il *trade-off* tra accuratezza della previsione e interpretabilità del modello
7. Misure di bontà di adattamento
 - Capacità predittiva "*in-sample*" e "*out-sample*"
 - Il *trade-off* tra distorsione e *variance*
 - La minimizzazione dell'errore quadratico medio
 - *Training-error* vs. *test-error*
 - I criteri di informazione
 - La relazione tra *Machine Learning* ed intelligenza artificiale
 - *Super-learning* e apprendimento dinamico

SESSIONE II: METODI DI RICAMPIONAMENTO E DI VALIDAZIONE

1. Stima del *test-error*
2. Metodi di validazione
 - Approccio con "insieme di validazione"
 - *K-fold cross-validation*
 - Approccio "*leave-one-out*"
3. Metodo *bootstrap*
4. L'algoritmo di *bootstrap*
5. *Bootstrap* vs. *cross-validation* ai fini della valutazione
6. Implementazione in Stata

SESSIONE III: SELEZIONE DEL MODELLO E SUA REGOLARIZZAZIONE

1. Selezione del modello e corretta specificazione
2. L'approccio con criteri di informazione
 - *Best subset selection*
 - *Backward stepwise selection*
 - *Forward stepwise Selection*
3. Metodi di regressione "*shrinkage*"
 - Regressione Lasso, Ridge ed elastica
 - Il Lasso adattivo
 - Criteri di informazione e *cross-validation* per il Lasso
4. Implementazione in Stata

<https://www.tstat.it/formazione/intro-machine-learning-stata/>



MACHINE LEARNING IN STATA: UN'INTRODUZIONE

SESSIONE IV: ANALISI DISCRIMINANTE E CLASSIFICATORE NEAREST- NEIGHBOR

1. Classificatore ottimale Bayesiano e frontiera decisionale
2. Tasso di errore di classificazione
3. Analisi discriminante
4. Analisi discriminante lineare e quadratica
5. Il classificatore *Naive-Bayes*
6. Il classificatore *k-nearest-neighbor*
7. Implementazione in Stata

SESSIONE V: REGRESSIONE NON- PARAMETRICA

1. Oltre i modelli parametrici: una panoramica
2. Approcci locali, semi-globali e globali
3. Metodi locali
 - Regressione *kernel*
 - Regressione *nearest-neighbor*
4. Metodi semi-globali
 - Funzione a gradini costante
 - *Piecewise polynomials*
 - Regressione *splineg*

SESSIONE VI: ALBERI DECISIONALI E METODI DI REGRESSIONE "ENSEMBLE"

1. Alberi di regressione e di classificazione
2. Costruire un albero decisionale tramite la "divisione binaria ricorsiva"
3. Potatura ottimale di un albero tramite *cross-validation*
4. Metodi di stima "ensemble" basati su alberi decisionali
 - *Bagging, Random Forests e Boosting*
5. Implementazione in Stata

LETTURE CONSIGLIATE

Econometric Analysis of Cross Section and Panel Data, Wooldridge, MIT Press (2010)

Microeconometrics Using Stata, Cameron e Trivedi, Revised Edition, StataPress (2010)

The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Hastie, T., Tibshirani, R., Friedman, J., Springer (2009)

An Introduction to Statistical Learning, Gareth, J., Witten, D., Hastie, T., Tibshirani, R., Springer (2013)

<https://www.tstat.it/formazione/intro-machine-learning-stata/>



MACHINE LEARNING IN STATA: UN'INTRODUZIONE

QUOTA DI ISCRIZIONE

La partecipazione al workshop è soggetta al pagamento della seguente quota di iscrizione:

Studente*: € 700.00
Accademico: € 950.00
Commerciale: € 1250.00

*Per usufruire dello status "studente" è necessario presentare copia del libretto universitario o un certificato di iscrizione (in carta semplice) all'Università ed essere *studenti a tempo pieno*. Studenti lavoratori dovranno considerare la tariffa Accademica.

I prezzi si intendono IVA 22% esclusa. L'aliquota IVA non sarà applicata per Enti Pubblici soggetti ad esenzione a norma dell'art. 14 c. 10 della L. 537/93 per la partecipazione a corsi di formazione dei propri dipendenti.

La quota di iscrizione include il pranzo, il materiale didattico e una licenza temporanea del software Stata (si consiglia di venire muniti del proprio computer o di chiedere informazioni alla segreteria per l'eventuale noleggio, al momento dell'iscrizione). Dà inoltre diritto ad uno sconto sull'acquisto di una nuova licenza per singolo utente del Software Stata (ad esclusione della versione per Studenti e del programma Grad Plan).

L'iscrizione al corso dovrà avvenire tramite lo specifico modulo di registrazione e pervenire a TStat S.r.l. entro il **1 Settembre 2020**. Lo svolgimento è condizionato dal raggiungimento di un numero minimo di 8 partecipanti ed un numero massimo di 12.

CONTATTI

Monica Gianni

TStat S.r.l. | Via Rettangolo, 12-14
I-67039 Sulmona (AQ)
T. +39 0864 210101

TStat Training | Kleebergstraße, 8
D-60322 Frankfurt am Main

formazione@tstat.it

www.tstat.it
www.tstattraining.eu

Ulteriori informazioni sulla modalità di iscrizione, incluso termini e condizioni di partecipazione sono disponibili nel nostro sito alla pagina <https://www.tstat.it/formazione/intro-machine-learning-stata/>.

