

# ASML: Primo laboratorio

Errore di adattamento e errore di generalizzazione

---

N. Torelli

Ottobre 2024

Università di Trieste - Dipartimento di Scienze Economiche Aziendali Matematiche e Statistiche "Bruno de Finetti"

## Prima parte

Scopo del laboratorio è quello di verificare il comportamento di una misura di adattamento in fase di apprendimento e in fase di valutazione

1. Si considerino i dati contenuti nel dataframe *mcycle* nel package *MASS* di R
2. Si suddivida casualmente il dataset in due parti: l'insieme di apprendimento (training) e quello di valutazione (Test set)
3. Si adatti al primo insieme, dapprima un modello lineare per la variabile *accel* usando come predittore *times*, poi si passi a un modello parabolico e infine cubico usando la funzione *lm*
4. Si usi ora per gli stessi modelli la funzione *poly()* all'interno di *lm* passando poi a modelli con grado superiore (si verifichi cosa accade con rappresentazioni grafiche)
5. Si ottenga ora per diversi gradi della funzione *poly* la misura dell'adattamento sul training set (si usi la devianza media)
6. Si valuti infine la stessa misura ma sul *test set* per i diversi gradi dell funzione
7. Si riproduca su un grafico la misura dell'adattamento nei due insiemi (training e test) per diversi livelli di flessibilità (oovero al variare del grado della funzione)

## Seconda parte

1. Si generino i dati secondo un modello definito da una funzione lineare a tratti più una componente casuale
2. A tal fine si può utilizzare lo script reso disponibile su moodle (esso genera una spline lineare)
3. Si può variare il livello di complessità della funzione (variando i parametri della spline o eventualmente anche il numero di nodi)
4. Si svolga l'analogo esercizio della prima parte sui dati simulati
5. (facoltativo) Si presenti il grafico sul test set della varianza e del bias al crescere della complessità (grado del polinomio del modello)