

VALUTARE UN SISTEMA ~ (10:00 min)

Supp. di avere tre sistemi: \triangleright_A , \triangleright_B e \triangleright_C che risolvono lo STESSO task. Per ogni sistema abbiamo delle MISURE DI PERFORMANCE, μ_A , μ_B e μ_C .

Se $\mu_B > \mu_A$ vogliamo sapere se questo risultato sia solo frutto del CASO, oppure se il sistema B effettivamente si comporta meglio di A anche se cambiamo il TEST SET utilizzato?

OSS: SE MACHINE LEARNING si basa sul fatto che il dataset utilizzato è i.i.d. rimetto al mondo.

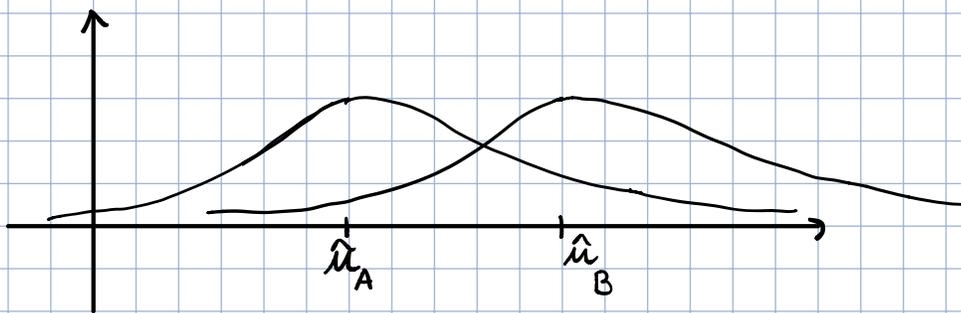
Per poter capire se i dati sono frutto del CASO dobbiamo stimare delle probabilità, e per avere delle buone stime devo avere PIU' OSSERVAZIONI. Per avere più osservazioni non DIVIDERE il TEST SET in modo RANDOMICO e SENZA OVERLAP. Possiamo replicare ogni sistema a ogni pezzo del TEST SET per ottenere una serie di osservazioni per ogni sistema.

METODO FREQUENTISTA := SE PARAMETRO da stimare
è un numero preciso.

METODO BAYESIANA := Possiamo definire una distr.
di mb. sul PARAMETRO
da stimare.

A questo punto non assumere che queste sequenze
di numeri non generate da una DISTRIBUZIONE
GAUSSIANA $N(\mu, \sigma^2)$. Tramite le osservazioni
non quindi in grado di STIMARE medie e
varianze per ogni sistema, ottenendo

$$\hat{\mu}_A, \hat{\sigma}_A^2, \hat{\mu}_B, \hat{\sigma}_B^2, \hat{\mu}_C, \hat{\sigma}_C^2$$



Per capire se le due distr. non separate non
utilizzare dei TEST STATISTICI che molto meno
confrontano elementi per elemento delle osservazioni
dei due sistemi. Non basta vedere se medie e
varianze si SOVRAPPONGONO.

TEST STATISTICI ~ (47:00 min)

Andiamo adesso a vedere in dettaglio alcuni test statistici.

SIGN TEST

Il SIGN TEST viene utilizzato per verificare se la diff. tra due v.a. X e Y abbia medie 0 o aumenti di X e Y non v.e. CONTINUE e che siano in grado di ottenere COPPIE di osservazioni da X e Y .

In particolare, posto $\pi := P(X > Y)$, il SIGN TEST assume come NULL HYPOTHESIS $H_0: \pi = 0.5$.

Per TESTARE H_0 si utilizzano le osservazioni $\{(x_1, y_1), \dots, (x_m, y_m)\}$ come segue:

- i) Si calcola il # W di coppie per cui $y_i - x_i > 0$.
- ii) Si confronta W con una $\text{Bin}(m, 0.5)$, dove m è il # di coppie che rimangono togliendo quelle in cui $x_i = y_i$.

Infatti, se H_0 è vera, allora $W \sim \text{Bin}(m, 0.5)$

Il SIGN TEST è un test NON-PARAMETRICO, in quanto non assume nulla sulle distribuzioni di X e Y .

Il nostro compito è quindi far vedere che, se H_0 è vera, la prob. di ottenere W è MOLTO BASSA, e quindi è insostenibile H_0 e viene RIGETTATA.

WILKINSON SIGN TEST

TODO

STUDENT T-TEST

Assume che le distr. X e Y sono GAUSSIANE, e quindi permette di ottenere risultati più forti.

Dati, X_1, \dots, X_m i.i.d., $X_i \sim N(\mu_1, \sigma^2)$

Y_1, \dots, Y_m i.i.d., $Y_i \sim N(\mu_2, \sigma^2)$

Calcolo

$$T := \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S^2}{m_1} + \frac{S^2}{m_2}}} \sim t((m_1 - 1) + (m_2 - 1))$$

dove S è detta POOLED VARIANCE ed è calcolata

come segue
$$S^2 = \frac{(m_1 - 1) \cdot S_1^2 + (m_2 - 1) \cdot S_2^2}{(m_1 - 1) + (m_2 - 1)}$$

ORACOLI PER VALUTARE I SISTEMI ~ (1:10:00 min)

Con l'avvento dei BIG DATA non stulti introdotti dei metodi alternativi per aumentare il # di osservazioni e per STACCARSI dai metodi di valutazione che utilizzano i TEST STATISTICI.

Questo nuovo metodo fa utilizzo degli ORACOLI.

SISTEMA S	SISTEMA X	ORACOLO
S_1	X_1	O_1
S_2	X_2	O_2

Un modo alternativo per capire se S è un "sistema buono" è quello di GENERARE tanti SISTEMI FITTIZI X e misurare le prob. che un sistema fittizio X performi meglio del sistema S.

Per generare questi sistemi fittizi NON PERMUTARE i risultati del sistema S.

Questo metodo può essere applicato quando non riesce ad applicare i TEST STATISTICI più famosi.

POS TAGGING (cont.)

~ (1:20:00 min)

Andiamo adesso a descrivere quali sono state le CLASSI GRAMMATICALI nelle due modelli sviluppati dai BIG del mondo del NLP.

§ BIG del mondo che andiamo a vedere sono:

- UNIVERSAL DEPENDENCIES

"All" viene sostituito con "A e" in quanto non si sa se il nominativo nome è maschile o femminile. Queste scelte è state fatte per SPECIALIZZARSI sulle parole e non guardare il contesto in cui la parola si trova o fine di RISPARMIARE del tempo di processamento.

- PENN TREEBANK

Q: In "to" utilizzato come PREPOSIZIONE viene categorizzato come un "to" o come un "in"?