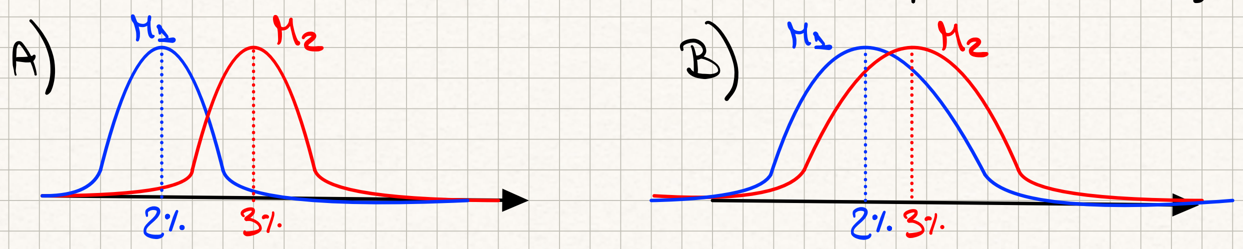
t-TEST, TEST PARAMETRICO STATISTICO

Test statistico utile a confrontare le metriche di confronto fra due modelli M_1 e M_2 (Per esempio per confrontare se le accuratezze rilevate siano simili oppure no).

Perché serve questo test?

Se noi consideriamo la distribuzione di probabilità dell'errore dei due modelli M_1 e M_2.
Supponiamo due situazioni dove $err(M_1) = 2\%$ e $err(M_2) = 3\%$.



Nel caso A, le due distribuzioni sono per lo più distinte, invece nel caso B sono quasi uguali.
Bisogna capire statisticamente se due risultati sono idendici o distinti.
Vogliamo vedere se la differenza fra le metriche di valutazione è significativa dal punto di vista statistico.

PASSI

1) Null-Hypotesis.

Presi i gruppi di risultati in cross-validation per M_1 e M_2, (Per esempio ogni modello ha un insieme di K valori di accuratezza), assumiamo che la distribuzione dei valori sia normale per ogni gruppo e che la varianza per ogni gruppo è simile.

L'ipotesi nulla è che la due distribuzioni sono identiche.

2) t-TEST. Confronta la differenza della media di due distribuzioni, in questo contesto useremo l'errore.
Valutiamo adesso l'errore $err(M_1)_i$ e $err(M_2)_i$ dove prendiamo l'i-esima divisione della cross-validation per entrambi i modelli. Indichiamo con $\overline{err}(M_1)$ e $\overline{err}(M_2)$ l'errore medio in cross-validation

$$t = \frac{\overline{err}(M_1) - \overline{err}(M_2)}{\sqrt{\frac{var(M_1 - M_2)}{K}}}$$

dove:

$$var(M_1 - M_2) = \frac{1}{K}\sum_{i=1}^{K}\left[err(M_1)_i - err(M_2)_i - \left(\overline{err}(M_1) - \overline{err}(M_2)\right)\right]^2$$

Questo test computa il t-statistic con K-1 gradi di libertà.
Per capire se i due risultati sono statisticamente diversi bisogna scegliere un valore significativo, tipicamente del 5% o 1%, questo valore è la probabilità che l'ipotesi nulla sia valida, consultare poi la tabella della distribuzion t sotto la voce (K-1) gradi di libertà.

PASSI

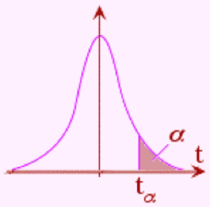
1) Calcolare t

2) Scegliere un livello significativo sig.
Ovvero decidiamo che con probabilità sig, $H_\emptyset$ si verifichi. Infatti tipicamente sig $\cong$ 5% o 1%.

3) Consultare la tabella per $(k-1)$ gradi di libertà.

4) $z = sig/2$ } LIMITE DI CONFIDENZA
Ovvero che nella tabella cerchiamo i valori di t corrispondenti a $(k-1)$ gradi di libertà e probabilità (α) di $sig/2$.

5) $t > z$ o $t < -z \Rightarrow$ Rigetta $H_\emptyset$.



T-Distribution Table

df	α = 0.1	0.05	0.025	0.01	0.005	0.001	0.0005
∞	t_α=1.282	1.645	1.960	2.326	2.576	3.091	3.291
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.920	4.303	6.965	9.925	22.328	31.600
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587