



Robot

Team Bravo



Presentazione problema

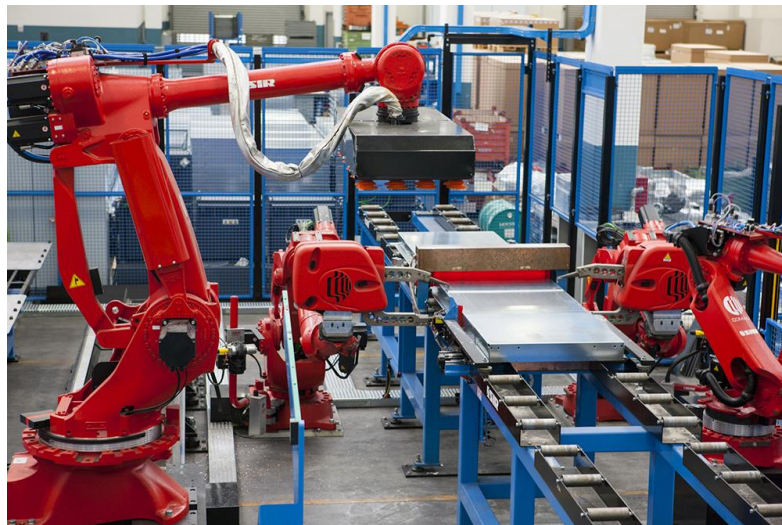
Il Problema del robot è una sfida che si pone l'obiettivo di riconoscere lo stato di funzionamento di un robot in base alle sue prestazioni.

Il problema si basa sui dati provenienti dai robot che eseguono cicli di lavorazione.

Consiste nel trovare un modo semplice e non supervisionato per distinguere i cicli in termini di stati di funzionamento e anomalie.



Robot da analizzare





Dataset

I dati rappresentano il funzionamento di un robot che svolge cicli di lavoro. Sono riportati i raw data di ciascun ciclo.

Le prime sei colonne riportano i valori di corrente, un valore per ciascuno dei sei bracci. Le successive sei colonne riportano i valori di posizione, un valore per ciascun braccio.

La colonna M contiene il tipo del di log

La colonna N contiene il timestamp

La colonna O contiene l'identificativo del ciclo di lavorazione

I dataset (cicli di lavorazione) sono organizzati in cinque gruppi e ogni gruppo di dati si trova nella rispettiva cartella: uno per i file del robot generati in condizioni nominali (nominal), gli altri tre per condizioni specifiche anomale, un quarto di test.



Dataset

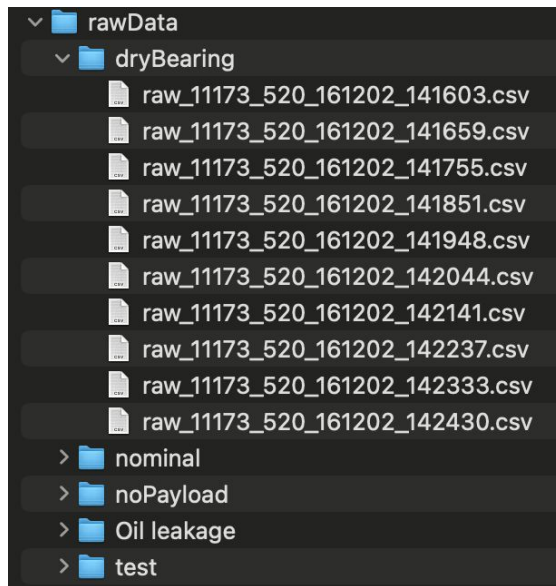
Le cartelle contenenti cicli di lavorazione anomali contengono le seguenti anomalie:

- Mancanza di olio (Oil Leakage)
- Cuscinetto a secco (Dry Bearing)
- Assenza di carico (No Payload)

Ogni dataset (ciclo di lavorazione) è in formato csv con punto e virgola come separatore di campo, punto come separatore decimale, senza intestazione.

Layout dei record (variabili, valori di esempio):

current arm 1: 0.1270;
current arm 2: 4.4355;
current arm 3: 3.3074;
current arm 4: 0.3184;
current arm 5: 0.0240;
current arm 6: 0.1242;
position arm 1: -43.8520;
position arm 2: -28.7582;
position arm 3: -55.5674;
position arm 4: -42.6073;
position arm 5: 65.3795;
position arm 6: 19.5901;
log type: 1;
timestamp: 2016-11-29 12:04:35.032;
processing cycle: 6





Record

I dati sono campionati per ogni ciclo di lavorazione ogni 2 millisecondi, i tempi non sono continui e quindi troveremo che tra i campionamenti di un ciclo di lavorazione e il suo successivo possano passare anche ore.

	Current_arm1	Current_arm2	Current_arm3	Current_arm4	Current_arm5	Current_arm6	Position_arm1	Position_arm2	Position_arm3	Position_arm4	Position_arm5	Position_arm6	Log_type	timestamp	Working_cycle
0	-0.03260	13.66200	6.89170	0.61060	-0.23850	-0.03140	-44.17770	-28.55290	-55.20300	-42.78060	65.61590	19.82230	1	2016-11-20 11:57:51.000	15
1	-0.00340	13.58210	6.89160	0.61060	-0.27270	-0.03160	-44.17770	-28.55280	-55.20300	-42.78060	65.61590	19.82230	1	2016-11-20 11:57:51.002	15
2	0.04160	13.65010	6.89160	0.61060	-0.27390	-0.01380	-44.17770	-28.55280	-55.20300	-42.78060	65.61590	19.82230	1	2016-11-20 11:57:51.004	15
3	-0.03140	13.63030	6.89160	0.54000	-0.25500	-0.03150	-44.17770	-28.55280	-55.20300	-42.78060	65.61590	19.82230	1	2016-11-20 11:57:51.006	15
4	-0.06960	13.63880	6.89150	0.61100	-0.27390	-0.04210	-44.17770	-28.55280	-55.20300	-42.78060	65.61590	19.82230	1	2016-11-20 11:57:51.008	15



Features

Le diverse features presenti nel dataset, che descrivono le correnti e le posizioni delle 6 braccia del robot e dati riguardo al ciclo di lavorazione attuale, sono:

- **Current_arm1, Current_arm2, Current_arm3, Current_arm4, Current_arm5, Current_arm6** : descrivono il valore della corrente in Ampere singolarmente per ogni braccio del robot.
- **Position_arm1, Position_arm2, Position_arm3, Position_arm4, Position_arm5, Position_arm6** : descrivono il valore della posizione in gradi singolarmente per ogni braccio.
- **Log_type** : descrive attraverso un numero identificativo con che tipo di log ci troviamo davanti
- **Timestamp** : descrive l'istante di tempo in cui è stato effettuato il campionamento
- **Working_cycle** : descrive con un numero identificativo a quale ciclo di lavorazione appartiene il campionamento



Preprocessamento dei dati

- si importano i dataset in questo modo: si concatenano i dataset di ogni cartella in singoli dataframe ottenendo così 4 dataframe ognuno per ogni caso specifico di lavorazione
- Si è appurato che i campionamenti siano effettuati in maniera continua per ogni ciclo di lavorazione senza lacune
- Si è appurato che non ci siano valori nulli (nan) all'interno dei dataset



Preprocessamento dei dati

Estrazione nuove features

Per adesso i dataframe ottenuti contengono tutte le righe di ogni ciclo di lavorazione, quindi per poter identificare ogni ciclo si è deciso di ipotizzare le features più importanti per ogni ciclo di lavorazione in maniera tale da poterlo rappresentare con una sola riga e quindi in uno spazio.

Per features più importanti si intendono quelle features che permettano una netta distinzione tra cicli di lavorazione di diverso tipo (nominali e anomali).

Quindi per riconoscere le features più importanti si è deciso di confrontare i dati statistici ricavati dai cicli di lavorazione di diverso tipo.



Preprocessamento dei dati

Estrazione nuove features

i dati statistici confrontati sono i seguenti :

- media per ogni colonna

	df	df1	df2	df3
Current_arm1	0.02318	0.07265	0.07299	0.06564
Current_arm2	6.88177	2.49607	2.85499	2.82578
Current_arm3	6.12424	2.07919	2.07670	2.07929
Current_arm4	0.19820	0.17355	0.17291	0.17645
Current_arm5	-1.02938	0.16114	0.16137	0.15549
Current_arm6	-0.22546	0.01066	0.01108	0.01028
Position_arm1	-11.42989	-11.09037	-11.03205	-11.13059
Position_arm2	-18.12738	-18.33179	-18.36498	-18.30689
Position_arm3	-56.47592	-56.83650	-56.75719	-56.88455
Position_arm4	-43.92329	-43.75049	-43.76357	-43.74140
Position_arm5	39.05559	38.81247	38.88054	38.76805
Position_arm6	17.38261	17.14871	17.14403	17.15055
Log_type	1.00000	1.00000	1.00000	1.00000
Working_cycle	11.52024	6.49783	6.51173	6.49841

df = nominal

df1 = Oil Leakage

df2 = No Payload

df3 = Dry bearing



Preprocessamento dei dati

Estrazione nuove features

- deviazione standard per ogni colonna

	df	df1	df2	df3
Current_arm1	4.11266	2.75751	2.76303	2.74863
Current_arm2	6.61873	6.28960	4.90750	4.88045
Current_arm3	4.25991	3.79888	3.80210	3.79272
Current_arm4	2.20881	1.01212	1.01363	1.22801
Current_arm5	0.92331	0.88830	0.89001	0.88477
Current_arm6	0.73673	0.62745	0.62873	0.62528
Position_arm1	21.85774	21.85393	21.83655	21.85368
Position_arm2	9.88144	9.88692	9.86957	9.89655
Position_arm3	22.95953	22.97184	22.96046	22.96991
Position_arm4	3.92431	3.94009	3.93847	3.93816
Position_arm5	39.93497	39.95974	39.93403	39.95505
Position_arm6	7.50549	7.52826	7.54201	7.50949
Log_type	0.00000	0.00000	0.00000	0.00000
Working_cycle	5.76838	2.86908	2.87102	2.86996

df = nominal
df1 = Oil Leakage
df2 = No Payload
df3 = Dry bearing



Preprocessamento dei dati

Estrazione nuove features

- valore minimo
per ogni colonna

	df	df1	df2	df3
Current_arm1	-20.80370	-12.66190	-12.62010	-12.69510
Current_arm2	-17.58740	-20.09080	-15.29590	-15.31340
Current_arm3	-18.00870	-17.04360	-17.00200	-17.00190
Current_arm4	-5.42730	-3.12530	-3.11590	-3.56700
Current_arm5	-4.00140	-2.50330	-2.50560	-2.50930
Current_arm6	-2.18730	-2.16380	-2.13520	-2.14210
Position_arm1	-44.17930	-43.85300	-43.85300	-43.85300
Position_arm2	-29.22850	-29.44680	-29.44670	-29.44670
Position_arm3	-101.65900	-102.02640	-102.02640	-102.02640
Position_arm4	-56.75800	-56.67750	-56.67760	-56.67760
Position_arm5	-31.46600	-31.70320	-31.70320	-31.70330
Position_arm6	11.64420	11.41200	11.41200	11.41200
Log_type	1.00000	1.00000	1.00000	1.00000
Working_cycle	2.00000	2.00000	2.00000	2.00000

df = nominal

df1 = Oil Leakage

df2 = No Payload

df3 = Dry bearing



Preprocessamento dei dati

Estrazione nuove features

- valore massimo per ogni colonna

	df	df1	df2	df3
Current_arm1	18.98250	12.81710	12.77550	12.81180
Current_arm2	32.83490	30.40710	25.99760	25.99870
Current_arm3	23.70510	17.87250	17.82140	17.81420
Current_arm4	9.72770	3.87160	3.90290	4.35220
Current_arm5	2.51620	2.56110	2.56080	2.55930
Current_arm6	2.54340	1.79860	1.79990	1.80190
Position_arm1	6.63080	6.95750	6.95750	6.95750
Position_arm2	4.06280	3.86320	3.86020	3.86020
Position_arm3	-34.29690	-34.66100	-34.66100	-34.66090
Position_arm4	-37.23480	-37.06060	-37.06070	-37.05940
Position_arm5	69.33770	69.10110	69.10120	69.10150
Position_arm6	41.67990	41.61680	41.61680	41.61680
Log_type	1.00000	1.00000	1.00000	1.00000
Working_cycle	21.00000	11.00000	11.00000	11.00000

df = nominal

df1 = Oil Leakage

df2 = No Payload

df3 = Dry bearing



Preprocessamento dei dati

Estrazione nuove features

- range per ogni colonna
(valore max - valore min)

	df	df1	df2	df3
Current_arm1	39.78620	25.47900	25.39560	25.50690
Current_arm2	50.42230	50.49790	41.29350	41.31210
Current_arm3	41.71380	34.91610	34.82340	34.81610
Current_arm4	15.15500	6.99690	7.01880	7.91920
Current_arm5	6.51760	5.06440	5.06640	5.06860
Current_arm6	4.73070	3.96240	3.93510	3.94400
Position_arm1	50.81010	50.81050	50.81050	50.81050
Position_arm2	33.29130	33.31000	33.30690	33.30690
Position_arm3	67.36210	67.36540	67.36540	67.36550
Position_arm4	19.52320	19.61690	19.61690	19.61820
Position_arm5	100.80370	100.80430	100.80440	100.80480
Position_arm6	30.03570	30.20480	30.20480	30.20480
Log_type	0.00000	0.00000	0.00000	0.00000
Working_cycle	19.00000	9.00000	9.00000	9.00000

df = nominal

df1 = Oil Leakage

df2 = No Payload

df3 = Dry bearing



Preprocessamento dei dati

Estrazione nuove features

- mediana per ogni colonna

	df	df1	df2	df3
Current_arm1	-0.17520	-0.01690	-0.01490	-0.02360
Current_arm2	9.21710	4.79495	3.67170	3.60430
Current_arm3	5.25040	0.59290	0.59300	0.58280
Current_arm4	-0.77580	0.18760	0.18750	0.13940
Current_arm5	-1.12710	0.22140	0.22240	0.19710
Current_arm6	-0.53360	0.01470	0.01620	0.01350
Position_arm1	6.33530	6.66100	6.66100	6.66100
Position_arm2	-25.86855	-26.12995	-26.18430	-26.05270
Position_arm3	-55.79360	-56.15960	-56.10950	-56.20060
Position_arm4	-44.15100	-43.97770	-43.97770	-43.97770
Position_arm5	62.90430	62.68710	62.76900	62.61410
Position_arm6	14.92490	14.69580	14.67670	14.71140
Log_type	1.00000	1.00000	1.00000	1.00000
Working_cycle	12.00000	6.00000	7.00000	7.00000

df = nominal
df1 = Oil Leakage
df2 = No Payload
df3 = Dry bearing



Preprocessamento dei dati

Estrazione nuove features

- Kurtosis per ogni colonna

	df	df1	df2	df3
Current_arm1	8.91001	8.45458	8.40813	8.52651
Current_arm2	1.80087	2.44489	3.35606	3.42408
Current_arm3	7.99727	4.98144	4.97760	4.97425
Current_arm4	1.80712	1.29854	1.28646	0.57354
Current_arm5	2.30882	1.44846	1.43096	1.47892
Current_arm6	0.62900	0.61719	0.60262	0.64117
Position_arm1	-1.65444	-1.65339	-1.64816	-1.65563
Position_arm2	-0.70766	-0.71080	-0.70288	-0.71335
Position_arm3	-0.70569	-0.70599	-0.70085	-0.70599
Position_arm4	1.60586	1.63246	1.63523	1.63236
Position_arm5	-1.03011	-1.03181	-1.02663	-1.03353
Position_arm6	1.63204	1.68287	1.67066	1.69672
Log_type	0.00000	0.00000	0.00000	0.00000
Working_cycle	-1.20668	-1.22106	-1.22314	-1.22310

df = nominal
df1 = Oil Leakage
df2 = No Payload
df3 = Dry bearing



Preprocessamento dei dati

Estrazione nuove features

- Skewness per ogni colonna

	df	df1	df2	df3
Current_arm1	0.50538	0.45230	0.45068	0.46134
Current_arm2	-0.16499	-0.16510	-0.01864	-0.00585
Current_arm3	-0.51930	-0.08594	-0.08334	-0.08507
Current_arm4	0.94101	-0.24024	-0.23903	-0.18854
Current_arm5	0.59251	-0.51422	-0.51391	-0.49875
Current_arm6	0.75533	-0.09067	-0.09268	-0.08979
Position_arm1	-0.50413	-0.50525	-0.51043	-0.50159
Position_arm2	0.74335	0.74278	0.74783	0.73979
Position_arm3	-0.72042	-0.72082	-0.72471	-0.71922
Position_arm4	-0.96346	-0.97519	-0.97626	-0.97149
Position_arm5	-0.87446	-0.87394	-0.87669	-0.87209
Position_arm6	1.54847	1.56207	1.56103	1.56240
Log_type	0.00000	0.00000	0.00000	0.00000
Working_cycle	-0.00557	0.00080	-0.00579	0.00074

df = nominal
df1 = Oil Leakage
df2 = No Payload
df3 = Dry bearing



Preprocessamento dei dati

Estrazione nuove features

le features più importanti riscontrate sono state inizialmente 2:

- media delle colonne dedicate alle correnti delle braccia robotiche

	df	df1	df2	df3
Current_arm1	0.02318	0.07265	0.07299	0.06564
Current_arm2	6.88177	2.49607	2.85499	2.82578
Current_arm3	6.12424	2.07919	2.07670	2.07929
Current_arm4	0.19820	0.17355	0.17291	0.17645
Current_arm5	-1.02938	0.16114	0.16137	0.15549
Current_arm6	-0.22546	0.01066	0.01108	0.01028

- deviazione standard delle colonne dedicate alle correnti delle braccia robotiche

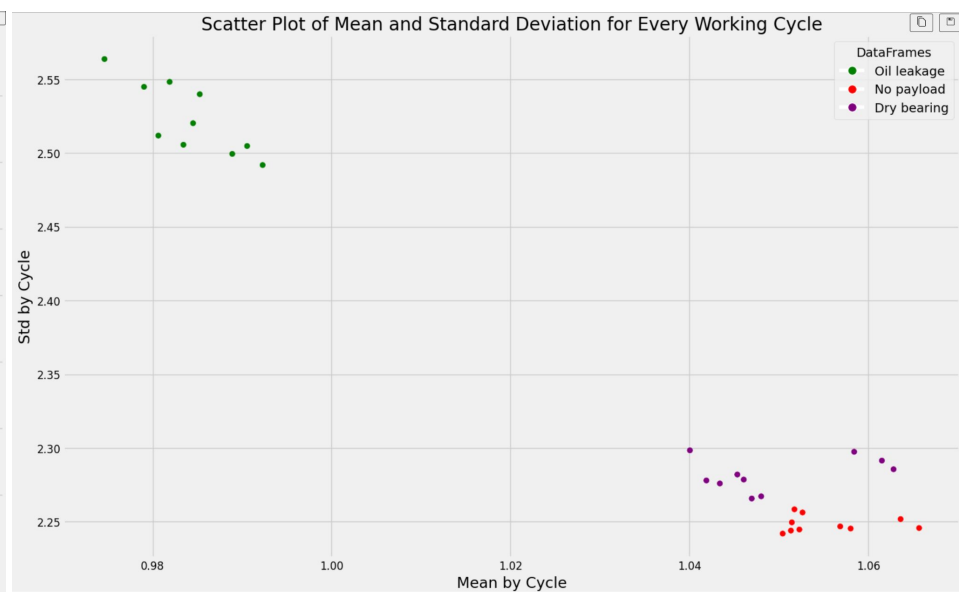
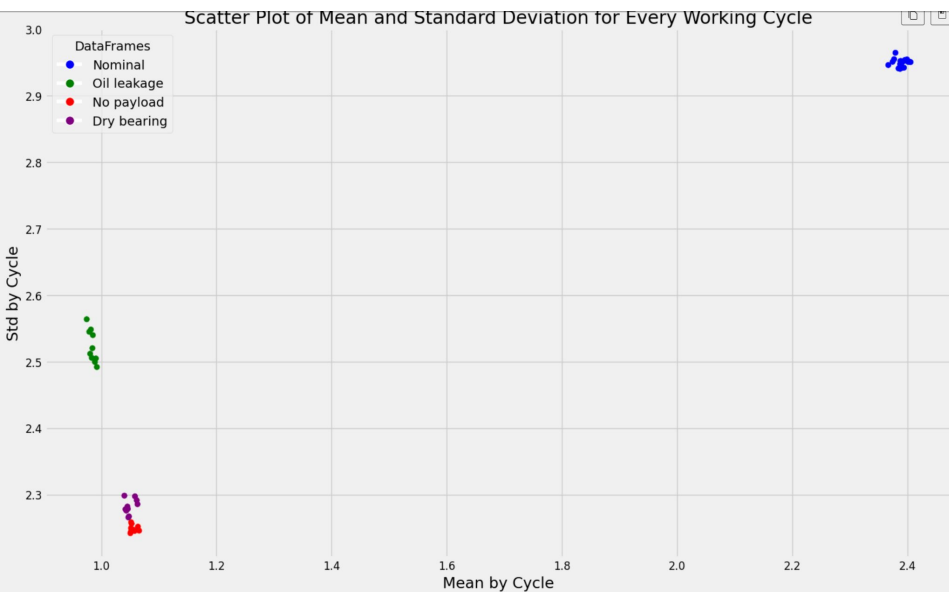
Current_arm1	4.11266	2.75751	2.76303	2.74863
Current_arm2	6.61873	6.28960	4.90750	4.88045
Current_arm3	4.25991	3.79888	3.80210	3.79272
Current_arm4	2.20881	1.01212	1.01363	1.22801
Current_arm5	0.92331	0.88830	0.89001	0.88477
Current_arm6	0.73673	0.62745	0.62873	0.62528



Preprocessamento dei dati

Estrazione nuove features

Estraendo le suddette features si è potuto visualizzare in uno spazio bidimensionale i cicli di lavorazione per poter vedere quanto fossero effettivamente raggruppabili in ottica di una clusterizzazione





Preprocessamento dei dati

Preparazione dei dataset X

Per il dataset destinato al primo training si è deciso di indicizzare in maniera diversa, in particolare si è deciso di creare per ogni riga un doppio indice, uno dedicato al tipo di stato di lavorazione e uno per il numero del ciclo di lavorazione, si ottiene quindi un dataframe con le due colonne delle nuove features (media delle correnti e deviazione standard) e doppio indice.

		Mean_by_Cycle	Std_by_Cycle
Working_cycle			
Nominal	2	2.40576	2.95054
	3	2.36707	2.94609
	4	2.39062	2.95001
Oil leakage	2	0.98344	2.50564
	3	0.99230	2.49187
	4	0.98190	2.54827
No payload	2	1.06361	2.25179
	3	1.05174	2.25842
	4	1.05145	2.24957
Dry bearing	2	1.04006	2.29842
	3	1.06152	2.29145
	4	1.04697	2.26574



Classificazione

Algoritmo

Per questo progetto è stato deciso di utilizzare come algoritmo il **K-means**. Il K-means spicca per la sua semplicità e robustezza, nel nostro caso permette di raggruppare in clusters insieme di cicli di lavorazione (divenuti punti in uno spazio) in base alla loro somiglianza (distanza).

Questo ci permetterà di riconoscere opportunamente i cicli di lavorazione di diversi stati nei corrispettivi clusters.

Una volta allenato il modello se gli viene fornito un nuovo ciclo di lavorazione quest'ultimo verrà associato al cluster più somigliante (vicino).



Classificazione

K-means

L'algoritmo K-Means è un metodo di clustering utilizzato per raggruppare dati in cluster basati sulla loro similarità. La sua logica si basa sull'idea di minimizzare la somma dei quadrati delle distanze tra i punti dati e i centroidi dei cluster a cui appartengono. Ecco come funziona l'algoritmo K-Means:

1. **Inizializzazione dei centroidi:** Si inizia scegliendo casualmente K (nel nostro caso 4, uno per ogni stato di lavoro del robot) centroidi iniziali, dove K è il numero di cluster desiderato
2. **Assegnazione dei punti ai cluster:** Per ogni punto dati, si calcola la distanza rispetto a ciascun centroide e si assegna il punto al cluster il cui centroide è più vicino.
3. **Aggiornamento dei centroidi:** Una volta che tutti i punti sono stati assegnati ai cluster, si ricalcolano i centroidi come la media di tutti i punti assegnati a ciascun cluster.
4. **Riassegnazione dei punti ai cluster:** Si ripete il passaggio 2, riassegnando i punti ai cluster in base ai nuovi centroidi calcolati.
5. **Convergenza:** Si ripetono i passaggi 3 e 4 fino a quando non si verifica una condizione di convergenza, ad esempio quando i centroidi non cambiano significativamente tra due iterazioni consecutive o quando il numero massimo di iterazioni è stato raggiunto.
6. **Output:** Una volta che l'algoritmo converge, restituisce i centroidi finali e l'assegnazione dei punti ai rispettivi cluster.



Classificazione

Training

il training viene quindi effettuato su un modello di clustering K-means con 4 clusters (nominale, no payload, dry bearing e oil leakage) e sul dataframe finale X al quale viene anche applicato una tecnica di standardizzazione dei valori delle features chiamata StandardScaler:

Lo StandardScaler è una tecnica di standardizzazione che normalizza le feature del dataset in modo che abbiano una media pari a 0 e una deviazione standard pari a 1. Questo processo di standardizzazione rende le feature comparabili tra loro, eliminando le differenze di scala che potrebbero influenzare negativamente l'efficacia degli algoritmi di apprendimento automatico

Motivazioni per l'uso dello StandardScaler:

1. **Sensibilità del K-Means alla scala:** Il K-Means misura la similarità tra i punti basandosi sulla distanza euclidea. Se le feature hanno scale diverse, alcune feature potrebbero influire maggiormente sulla distanza rispetto ad altre, portando a una clusterizzazione distorta.
2. **Equità tra le feature:** L'uso dello StandardScaler garantisce che ogni feature contribuisca in modo equo al processo di clustering. Questo è particolarmente importante quando le feature rappresentano quantità diverse o hanno unità di misura diverse.
3. **Miglioramento della convergenza:** La standardizzazione accelera la convergenza del K-Means. Dato che l'algoritmo minimizza la somma delle distanze quadrate, una scala uniforme facilita la convergenza dell'algoritmo.
4. **Miglior interpretazione:** La standardizzazione dei dati semplifica l'interpretazione dei risultati. I centroidi dei cluster saranno più rappresentativi e significativi quando le feature sono sulla stessa scala.



Classificazione

Training

Il processo di addestramento ha una durata media di circa 1 secondo, tuttavia, non sempre riesce a effettuare una clusterizzazione accurata degli stati, evidenziando un significativo errore nella distinzione tra "No Payload" e "Dry Bearing".

		Mean_by_Cycle	Std_by_Cycle	Cluster
Nominal	2	2.40576	2.95054	1
	3	2.36707	2.94609	1
	4	2.39062	2.95001	1
Oil leakage	2	0.98344	2.50564	2
	3	0.99230	2.49187	2
	4	0.98190	2.54827	2
No payload	2	1.06361	2.25179	0
	3	1.05174	2.25842	0
	4	1.05145	2.24957	0
Dry bearing	2	1.04006	2.29842	3
	3	1.06152	2.29145	3
	4	1.04697	2.26574	0

errata clusterizzazione

		Mean_by_Cycle	Std_by_Cycle	Cluster
Nominal	2	2.40576	2.95054	1
	3	2.36707	2.94609	1
	4	2.39062	2.95001	1
Oil leakage	2	0.98344	2.50564	2
	3	0.99230	2.49187	2
	4	0.98190	2.54827	2
No payload	2	1.06361	2.25179	0
	3	1.05174	2.25842	0
	4	1.05145	2.24957	0
Dry bearing	2	1.04006	2.29842	3
	3	1.06152	2.29145	3
	4	1.04697	2.26574	3

corretta clusterizzazione



Classificazione

Predizioni sul dataset di testing

il modello ottenuto nel caso di corretta clusterizzazione si è usato poi per provare a classificare i cicli di lavorazione forniti dal dataset di test ottenendo il seguente risultato

		Mean_by_Cycle	Std_by_Cycle	Predicted Cluster
Working_cycle				
Test	2	1.04905	2.26860	3
	3	1.04272	2.26373	3
	4	1.04921	2.24713	0
	5	1.05598	2.24342	0
	6	1.04802	2.26013	3
	7	1.05354	2.24250	0
	8	1.05308	2.26003	3
	9	1.04872	2.27317	3
	10	1.03749	2.29906	2
	11	1.03299	2.31825	2
	12	1.02374	2.34533	2
	13	1.03088	2.36725	2
	14	1.01011	2.39113	2
	15	0.99815	2.43084	2
	16	1.00333	2.44870	2

0 = No payload
1 = Nominal
2 = Oil Leakage
3 = Dry Bearing



Classificazione

Cross Validation

Per validare il modello, abbiamo adottato l'approccio della cross-validation. In un numero fissato di iterazioni, nel nostro caso 5, abbiamo estratto stratificatamente N righe dal dataset originale, dove N è stato impostato a 3. La stratificazione è stata eseguita considerando i diversi stati di lavorazione, garantendo così una rappresentazione equa di ciascuna categoria. Le righe estratte durante questo processo hanno costituito il dataset di validazione, mentre il dataset completo, privato delle righe estratte, è stato utilizzato come dataset di training. Successivamente, abbiamo addestrato il modello utilizzando il dataset di training e lo abbiamo testato sulla previsione sia del dataset di validazione che del dataset di testing. Il criterio di successo per la validità del modello è stato valutato attraverso la coerenza nella clusterizzazione del dataset di validazione durante ciascuna iterazione. Se il modello è in grado di clusterizzare correttamente il dataset di validazione e produce risultati simili sul dataset di testing in ogni iterazione, possiamo concludere di aver ottenuto un modello valido e affidabile.



Classificazione

Risultati Cross Validation

Per questo primo tentativo, i risultati sono stati accettabili, con un'accuratezza (media delle corrette clusterizzazioni del dataset di validazione) che variava tra 0,75 e 0,95. Tuttavia, la clusterizzazione del dataset di testing è risultata essere instabile, mostrando risultati diversi ad ogni iterazione. Di conseguenza, il modello si è rivelato poco affidabile e, pertanto, non valido.

```
Estrazione di n cicli dai singoli dataset:

Working_cycle
Nominal      18      1
             20      1
             16      1
Oil leakage   4       2
             11      2
             10      2
No payload    2       0
             9       0
             6       0
Dry bearing   9       3
             2       3
             11      3
Name: Cluster, dtype: int32

Assegnazione degli n cicli ai cluster:

array([1, 1, 1, 2, 2, 2, 3, 3, 3, 3, 3, 3], dtype=int32)

Previsione del dataset di testing:

array([3, 3, 3, 3, 3, 3, 1, 3, 2, 2, 2, 2, 2, 2], dtype=int32)
```

Esempio di risultati
della cross validation



Classificazione

Seconda Prova

Per migliorare le prestazioni del modello e conferirgli maggiore affidabilità, abbiamo optato per l'aggiunta di ulteriori feature per la rappresentazione di ogni ciclo di lavorazione. Attraverso una nuova analisi dei dati statistici per ciascun ciclo, siamo andati alla ricerca di features che potessero facilitare la differenziazione tra No Payload e Dry Bearing e abbiamo identificato e distinto le seguenti features:

- Range della colonna dedicata alla corrente del braccio 4

	Nominal	Oil leakage	No payload	Dry bearing
Current_arm2	50.422300	50.497900	41.293500	41.312100
Current_arm3	41.713800	34.916100	34.823400	34.816100
Current_arm4	15.155000	6.996900	7.018800	7.919200

- Skewness della colonna dedicata alla corrente del braccio 2

	Nominal	Oil leakage	No payload	Dry bearing
Current_arm2	-0.164995	-0.165104	-0.018643	-0.005848

- Kurtosis della colonna dedicata alla corrente del braccio 4

	Nominal	Oil leakage	No payload	Dry bearing
Current_arm2	1.800873	2.444886	3.356061	3.424080
Current_arm3	7.997271	4.981435	4.977601	4.974249
Current_arm4	1.807118	1.298541	1.286464	0.573538



Classificazione

Seconda Prova (training)

Una volta preparato il nuovo dataset con le nuove feature nella stessa maniera come il precedente si avvia il processo di addestramento e anche qui ha una durata media di circa 1 secondo, ma in questo caso riesce sempre a clusterizzare correttamente tutti i cicli di lavorazione.

		Mean_by_Cycle	Std_by_Cycle	Range_by_Cycle	Skewness_by_Cycle	Kurtosis_by_Cycle	Cluster
Working_cycle							
Nominal	2	2.40576	2.95054	15.06810	-0.18623	1.82601	0
	3	2.36707	2.94609	15.09010	-0.12657	1.86365	0
	4	2.39062	2.95001	14.97090	-0.16577	1.75931	0
Oil leakage	2	0.98344	2.50564	6.95920	-0.15308	1.28606	2
	3	0.99230	2.49187	6.98310	-0.15807	1.30754	2
	4	0.98190	2.54827	6.97110	-0.17502	1.29548	2
No payload	2	1.06361	2.25179	7.01880	-0.02574	1.27565	3
	3	1.05174	2.25842	7.00470	-0.01679	1.23929	3
	4	1.05145	2.24957	6.99140	-0.01258	1.28783	3
Dry bearing	2	1.04006	2.29842	7.91920	0.00129	0.37432	1
	3	1.06152	2.29145	7.80990	-0.01769	0.55249	1
	4	1.04697	2.26574	7.47540	-0.00116	0.73702	1



Classificazione

Seconda Prova (Predizioni sul dataset di testing)

il modello ottenuto è stato usato poi per prevedere la clusterizzazione dei cicli di lavorazione forniti dal dataset di test, ottenendo i seguenti risultati:

	Mean_by_Cycle	Std_by_Cycle	Range_by_Cycle	Skewness_by_Cycle	Kurtosis_by_Cycle	Predicted Cluster
Working_cycle						
Test 2	1.04905	2.26860	6.98260	-0.03111	1.30385	3
3	1.04272	2.26373	6.98690	-0.02035	1.33392	3
4	1.04921	2.24713	6.95210	-0.01583	1.30748	3
5	1.05598	2.24342	6.98020	-0.01980	1.30862	3
6	1.04802	2.26013	6.99140	-0.02012	1.27687	1
7	1.05354	2.24250	6.97630	-0.00793	1.34332	3
8	1.05308	2.26003	7.01060	-0.03652	1.30112	0
9	1.04872	2.27317	6.98630	-0.04493	1.32012	0
10	1.03749	2.29906	6.98240	-0.04392	1.29284	2
11	1.03299	2.31825	6.95050	-0.06598	1.29791	2
12	1.02374	2.34533	6.98630	-0.08117	1.28025	2
13	1.03088	2.36725	7.00430	-0.10879	1.28608	0
14	1.01011	2.39113	6.96120	-0.10911	1.29866	2
15	0.99815	2.43084	6.94680	-0.11730	1.25331	2
16	1.00333	2.44870	6.97190	-0.14066	1.28394	2

0 = Nominale
1 = Dry Bearing
2 = Oil Leakage
3 = No Payload



Classificazione

Seconda prova (Cross Validation)

La cross-validation viene eseguita seguendo la stessa procedura descritta precedentemente, ma in questo caso i risultati hanno confermato la completa validazione e affidabilità del modello. L'accuratezza ha costantemente raggiunto il massimo (1), e la clusterizzazione sul dataset di testing ha prodotto sempre lo stesso risultato ad ogni iterazione.

```
Working_cycle
Nominal      2      0
             15      0
             21      0
Oil leakage  4      2
             5      2
             7      2
No payload   2      3
             5      3
             7      3
Dry bearing  11     1
             4      1
             7      1
Name: Cluster, dtype: int32

array([0, 0, 0, 2, 2, 2, 3, 3, 3, 1, 1, 1], dtype=int32)

array([3, 0, 3, 3, 1, 3, 0, 0, 2, 2, 2, 0, 2, 2, 2], dtype=int32)

Average Accuracy: 1.0
```



Conclusioni

In conclusione, il presente progetto robot ha rappresentato un significativo passo avanti nella nostra comprensione e implementazione di progetti di machine learning. Attraverso l'analisi dettagliata dei cicli di lavorazione e l'applicazione di algoritmi di clustering, siamo riusciti a ottenere risultati promettenti.

Durante il primo tentativo, l'affidabilità del modello è risultata non sufficientemente stabile, evidenziando problematiche nella clusterizzazione del dataset di testing, in particolare per quanto riguarda la distinzione tra gli stati "No Payload" e "Dry Bearing". Tuttavia, la successiva ottimizzazione del modello mediante l'aggiunta di nuove features ha comportato un netto miglioramento delle prestazioni. La scelta di implementare una cross-validation ha dimostrato la solidità del modello, con un'accuratezza costantemente massima (1) e risultati coerenti sulla clusterizzazione del dataset di testing ad ogni iterazione.



Conclusioni

Possibili evoluzioni

Le possibili evoluzioni del progetto offrono un terreno fertile per ulteriori miglioramenti e raffinamenti del sistema robotico. Un approccio potenzialmente fruttuoso consiste nell'espandere il set di features utilizzato per la rappresentazione di ciascun ciclo di lavorazione.

Un'altra direzione promettente riguarda l'ottimizzazione dell'algoritmo di clustering. L'aggiunta di funzionalità avanzate, come l'inizializzazione ciclica dei centroidi, potrebbe contribuire a migliorare la convergenza dell'algoritmo, favorendo risultati più consistenti e riducendo la sensibilità agli stati iniziali dei centroidi. Questa strategia mira a ottenere una somma di distanze più bassa, indicatore di una clusterizzazione più precisa e stabile.

Inoltre, potrebbe essere interessante esplorare nuovi algoritmi di clustering che potrebbero offrire vantaggi specifici nel contesto del progetto robot. L'adozione di metodologie innovative potrebbe contribuire a superare eventuali limitazioni degli algoritmi attuali, portando a una maggiore precisione nella categorizzazione degli stati di lavorazione.



Conclusioni

Progetto Svolto dal Team Bravo

Manuele Longo

Davide Cologgi

Lorenzo Nicotera