Feature selection & Feature Extraction

Rimini – 15/11/2021

Valentina Pellicioni

Studente di Dottorato in «Scienza e Cultura del Benessere e degli Stili di Vita»

Dipartimento di Scienze per la Qualità della vita

Valentina Pellicioni - valentina.pellicion2@unibo.it

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
VARLAB: VIRTUAL AND AUGMENTED REALITY LAB

CV (in breve)

- Studentessa di dottorato in AI applicata al drug discovery
- Laurea magistrale a ciclo unico in Farmacia

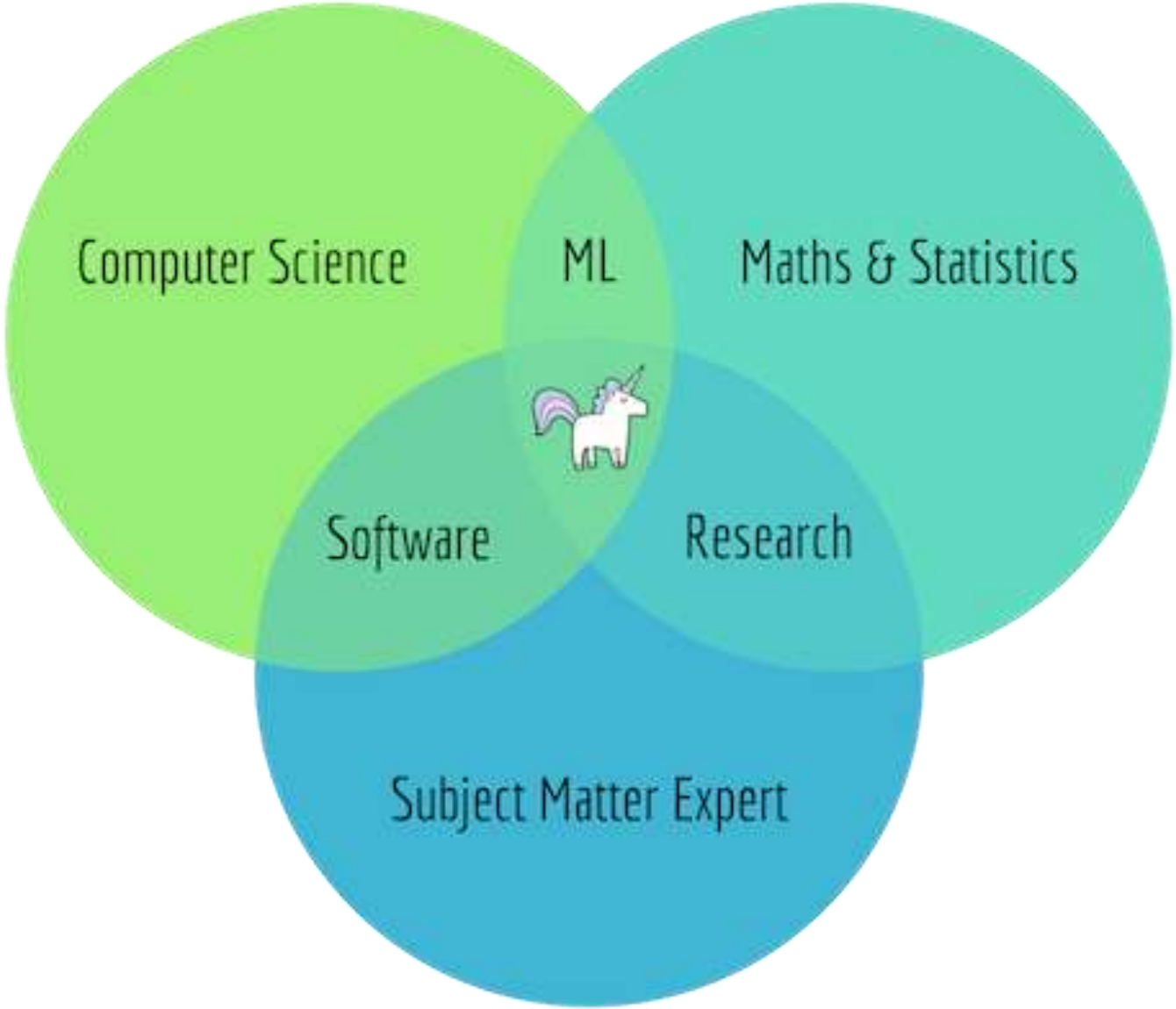
Contatti

Principali aree di ricerca

- Computer Aided Drug Discovery
- Machine Learning e Deep Learning

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DATA SCIENCE



Unicorni fantastici e dove trovarli (data science)

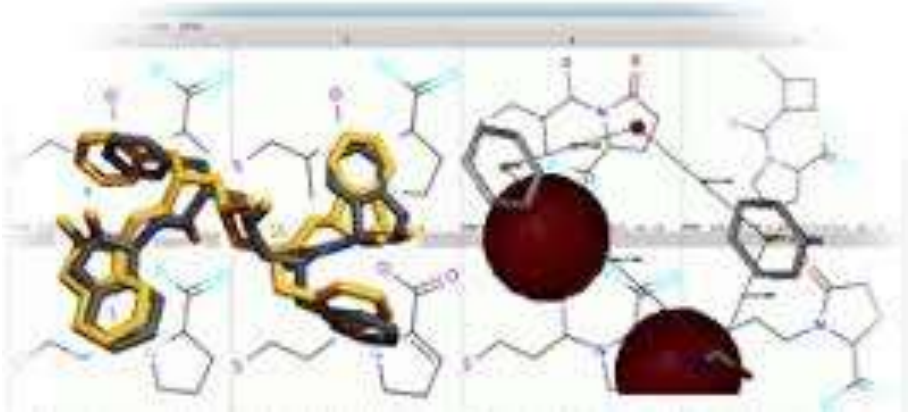
Ispirazione

Chimica e Farmacologia

Studio dei tumori

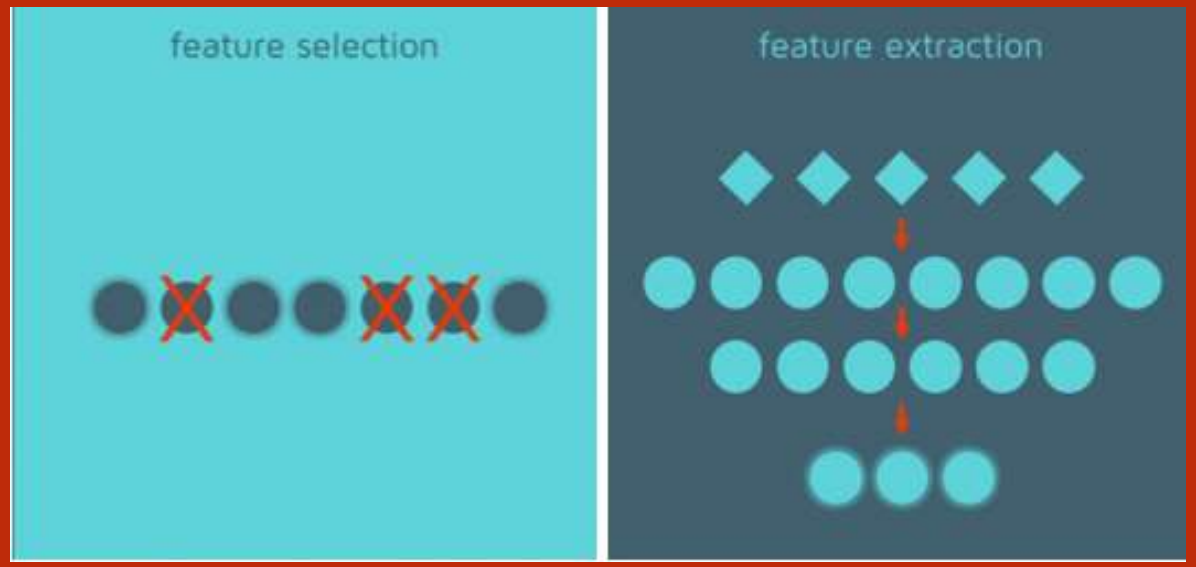
Machine and Deep Learning

Chemoinformatica



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Feature Selection & Feature Extraction - Introduzione



Valentina Pellicioni

Studente di Dottorato in «Scienza e Cultura del Benessere e degli Stili di Vita»

Dipartimento di Scienze per la Qualità della vita

Com'è fatto un dataset

Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3	1.4	0.2	Iris-setosa
3	4.7	3.2	1.3	0.2	Iris-setosa
4	4.6	3.1	1.5	0.2	Iris-setosa
5	5	3.6	1.4	0.2	Iris-setosa
6	5.4	3.9	1.7	0.4	Iris-setosa
7	4.6	3.4	1.4	0.3	Iris-setosa
8	5	3.4	1.5	0.2	Iris-setosa
9	4.4	2.9	1.4	0.2	Iris-setosa
10	4.9	3.1	1.5	0.1	Iris-setosa
11	5.4	3.7	1.5	0.2	Iris-setosa
12	4.8	3.4	1.6	0.2	Iris-setosa
13	4.8	3	1.4	0.1	Iris-setosa
14	4.3	3	1.1	0.1	Iris-setosa

Rows (righe): contengono i samples (campioni)

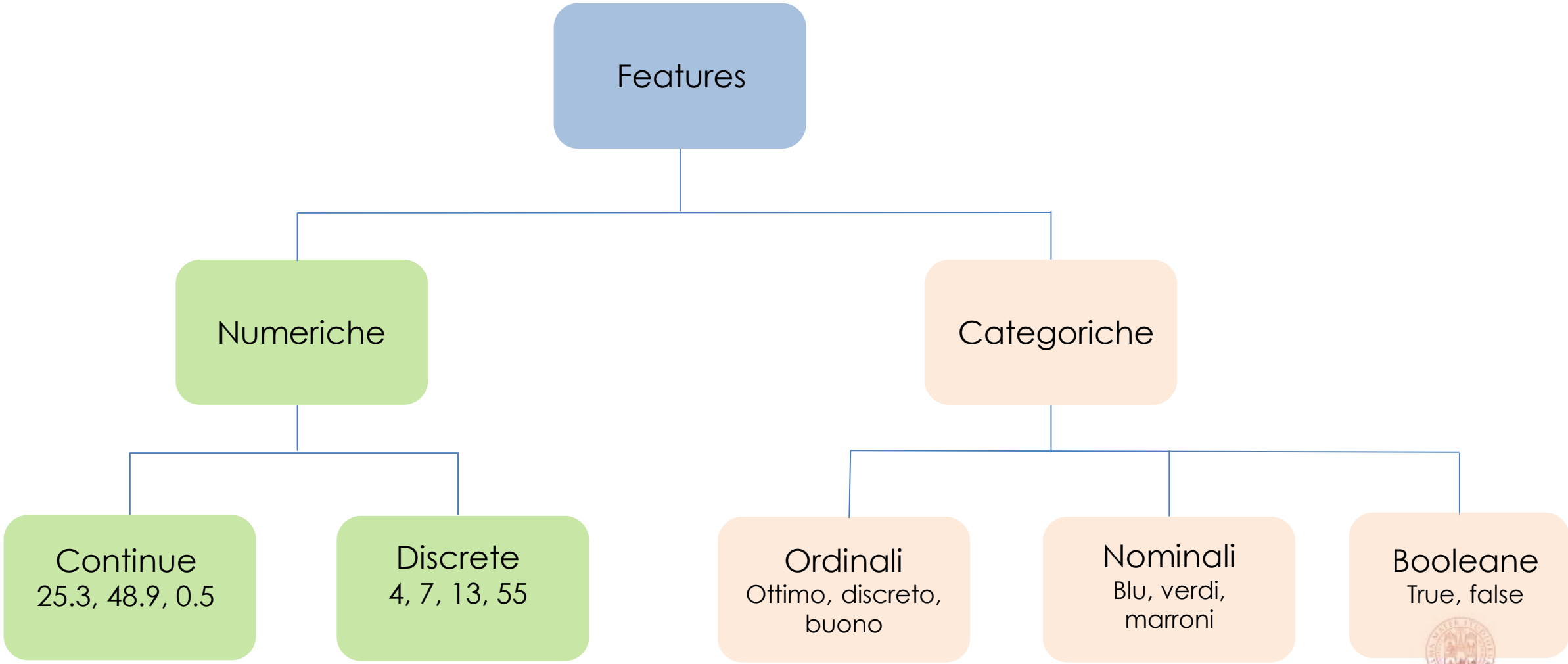
Columns (colonne): contengono le **features** (caratteristiche) o variabili o dimensioni

Che cosa sono le features?

- Una feature è una proprietà del campione che si sta cercando di analizzare che può essere misurata o contata.

- A seconda di ciò che state cercando di analizzare, le features che includete nel dataset possono variare ampiamente e, di conseguenza, varieranno gli algoritmi impiegati per il modello.

Come si presentano le features?



Dataset reale (Titanic)

	A	B	C	D	E	F	G	H	I	J	K	L
1	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
2	1	0	3	Braund, Mr. (	male	22	1	0	A/5 21171	7.25		S
3	2	1	1	Cumings, Mr	female	38	1	0	PC 17599	71.2833	C85	C
4	3	1	3	Heikkinen, M	female	26	0	0	STON/O2. 31	7.925		S
5	4	1	1	Futrelle, Mrs	female	35	1	0	113803	53.1	C123	S
6	5	0	3	Allen, Mr. Wi	male	35	0	0	373450	8.05		S
7	6	0	3	Moran, Mr. J	male		0	0	330877	8.4583		Q

Variabile categorica convertita

Variabile categorica nominale

Variabile numerica discreta

Variabile numerica continua

Perché le features sono importanti?

- Le features agiscono come un input nel sistema, infatti vengono usate dai modelli per fare previsioni.

- La qualità delle feature influisce direttamente sulla qualità delle intuizioni che è possibile ottenere dal modello

- Diversi problemi all'interno dello stesso settore possono richiedere feature diverse, perciò è importante avere una forte comprensione degli obiettivi del vostro progetto di data science.

Problematiche legate ad un numero elevato di features

Difficoltà ad analizzare i dati

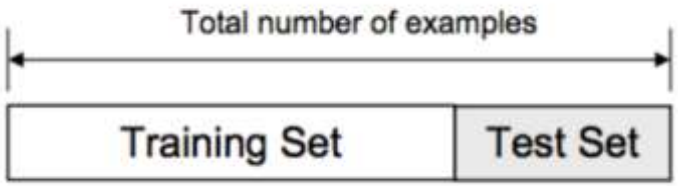
Difficoltà a visualizzare i dati

Difficoltà ad addestrare il modello → Curse of dimensionality

Features > 100

Curse of Dimensionality

La maledizione della dimensionalità sorge quando si lavora con dati altamente dimensionali. "Man mano che il numero di dimensioni cresce, la quantità di dati di cui abbiamo bisogno per generalizzare accuratamente cresce esponenzialmente".



La fase di test ci aiuta a stabilire se il modello è capace di **generalizzare** o no.

La generalizzazione è l'abilità del modello di predire correttamente il risultatro per un input non visto durante la fase di training. Se ciò non avviene si verificano **underfitting** o **overfitting**.

Segnale VS rumore

Se Il nostro modello **non generalizza bene** i dati di apprendimento ai dati di convalida, cioè nuovi dati che non sono stati utilizzati nella fase di training, le sue prestazioni saranno molto scarse.

Per generalizzare bene il modello addestrato deve saper riconoscere la differenza tra **segnale** e **rumore.**

Segnale: ciò che vogliamo imparare dai dati

Rumore: informazioni irrilevanti

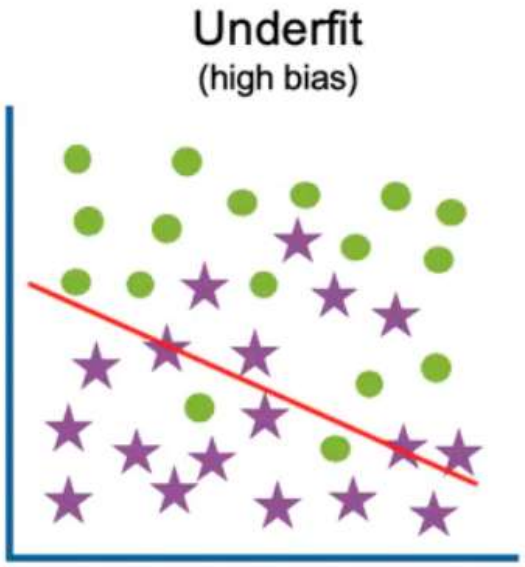
Bias e varianza

Per capire meglio perché un modello non riesce a generalizzare dobbiamo conoscere anche i concetti di **bias** e **varianza.**

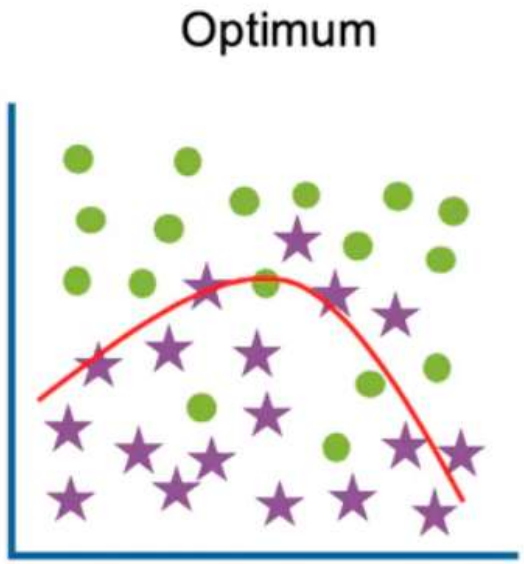
Bias: è la differenza tra la previsione media del nostro modello e il valore corretto che stiamo cercando di prevedere. Non è altro che l'**errore sistematico**, che non dipende dalla casualità dei dati.

Varianza: rappresenta la variabilità della previsione del modello, se si addestra il modello più volte su diversi dataset. Ci dice, cioè, quanto il modello è sensibile alla **casualità dei dati** nel set di addestramento.

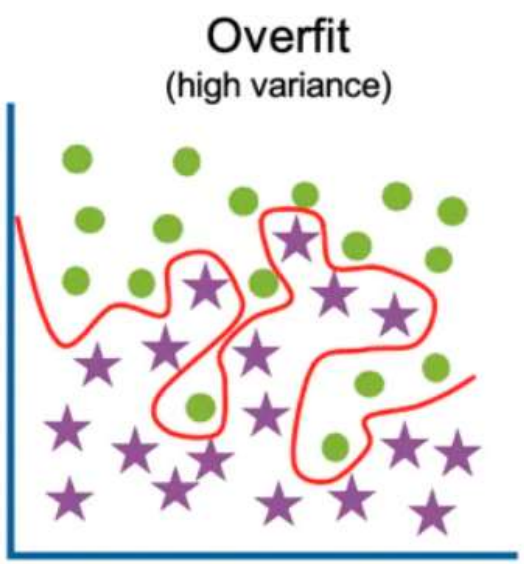
Underfitting e overfitting



High training error
High test error



Low training error
Low test error

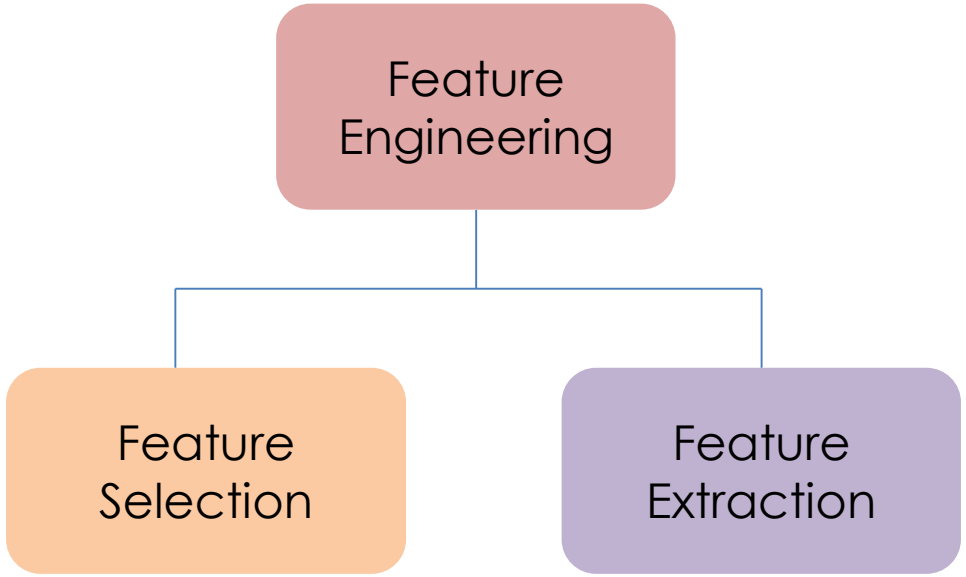


Low training error
High test error

Fighting the curse of dimensionality

La **riduzione della dimensionalità** è il processo di riduzione del numero di feature.

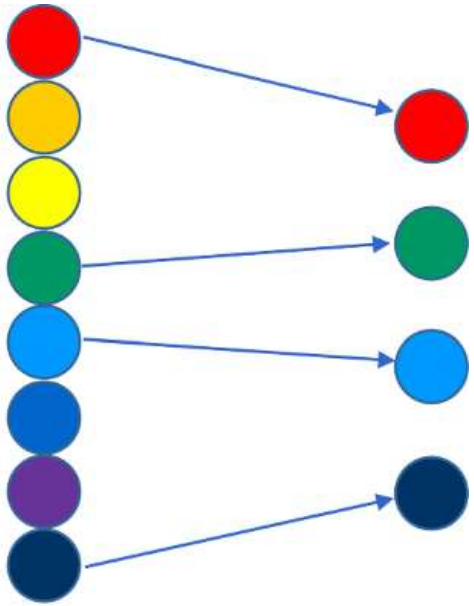
Viene realizzata attraverso una serie di tecniche complessivamente chiamate feature engeneering che si dividono in due categorie: feature selection e feature extraction.



Quali feature posso eliminare?

In base al loro ruolo ai fini della modellazione le feature possono essere:

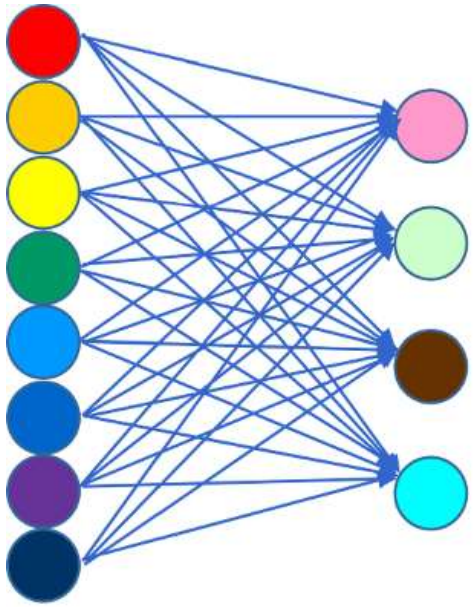
- **Rilevanti**: hanno un'influenza sull'output e sono indispensabili per la creazione di un modello capace di generalizzare correttamente.
- **Irrilevanti**: non hanno alcuna influenza sull'output, possiedono valori casuali per ogni dato.
- **Ridondanti**: Una ridondanza esiste ogni volta che due o più feature codificano lo stesso tipo di informazione sul dato.



Feature selection

Pro: le informazioni importanti relative a una singola feature sono mantenute

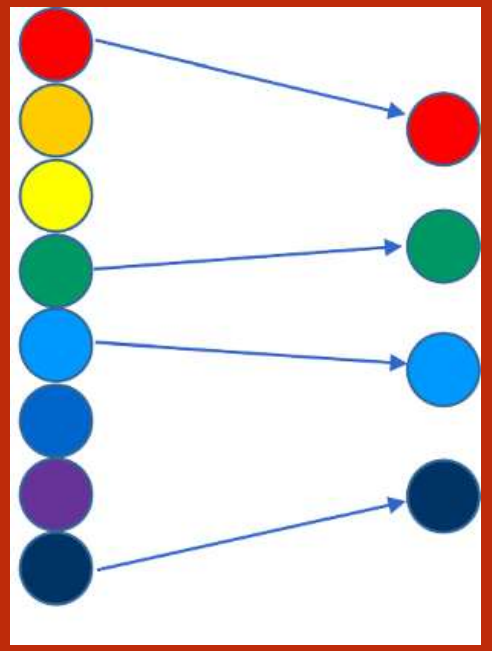
Contro: se vogliamo ottenere un piccolo insieme di feature e le feature originali sono molto diverse, c'è la possibilità di perdere informazioni poiché alcune delle feature devono essere omesse.



Feature extraction

Pro: la dimensione dello spazio delle feature può essere ridotta senza perdita di informazioni.

Contro: la combinazione lineare delle feature originali di solito non è interpretabile e l'informazione su quanto contribuisce una feature originale è spesso persa.



Feature Selection

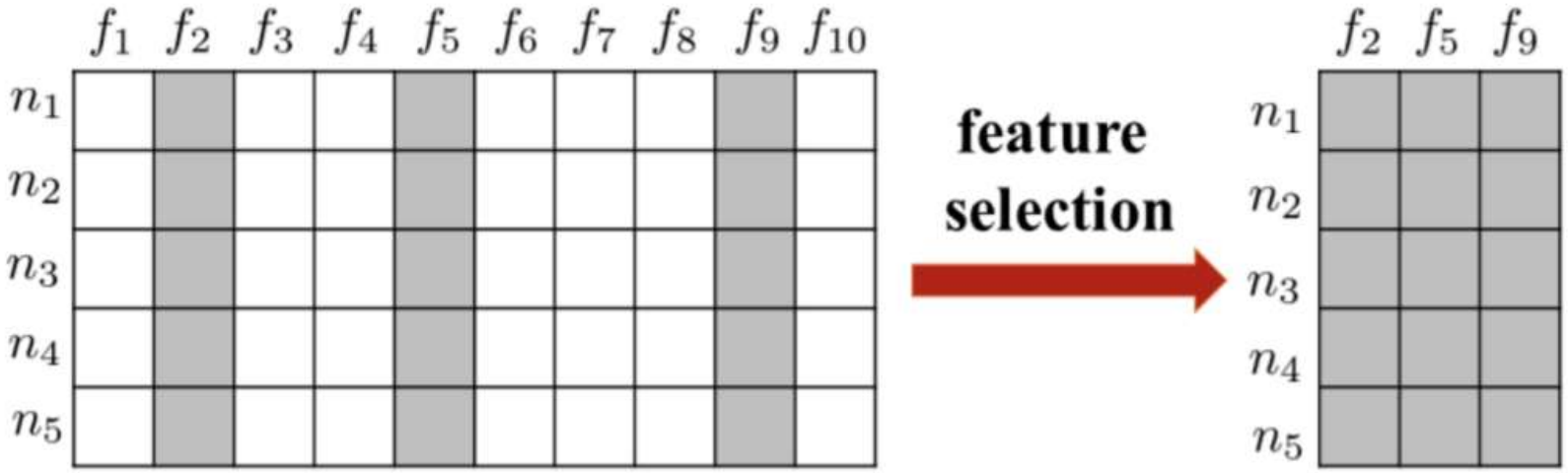
Valentina Pellicioni

Studente di Dottorato in «Scienza e Cultura del Benessere e degli Stili di Vita»

Dipartimento di Scienze per la Qualità della vita

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Riduzione della matrice dei dati



Il processo per trovare queste matrici ristrette è chiamato riduzione della dimensionalità.

Quali problemi è in grado di risolvere

1. Riduce la dimensionalità dello spazio delle feature: limita i requisiti di stoccaggio e aumenta la velocità dell'algoritmo. Riduce il problema della small sample size.

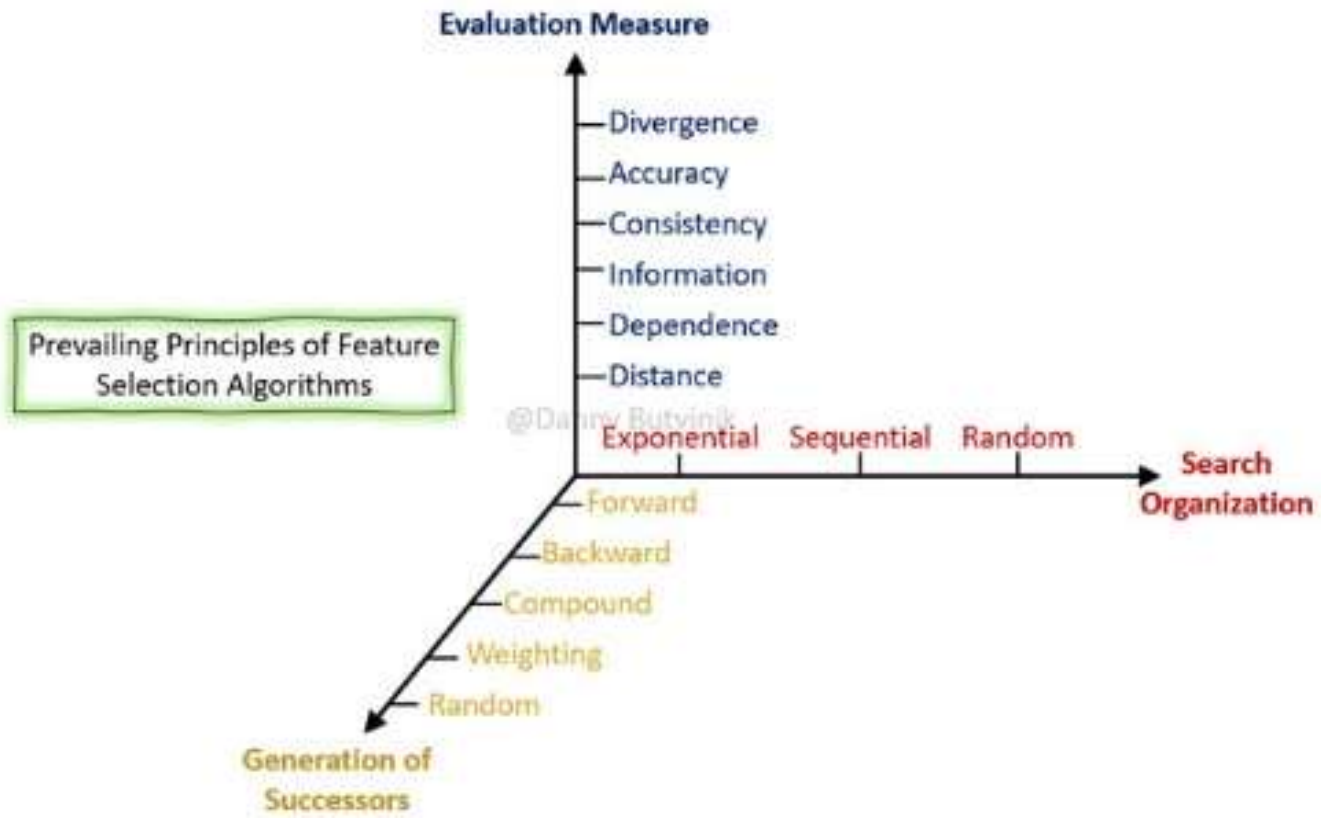
2. Rimuove i dati ridondanti, irrilevanti o rumorosi, migliorandola qualità dei dati. Di conseguenza riduce l'overfitting: con meno dati ridondanti diminuisce la probabilità di prendere decisioni basate sul rumore.

3. Migliora la precisione: con meno dati fuorvianti migliorano sia l'accuratezza della modellazione che la comprensibilità dei risultati.

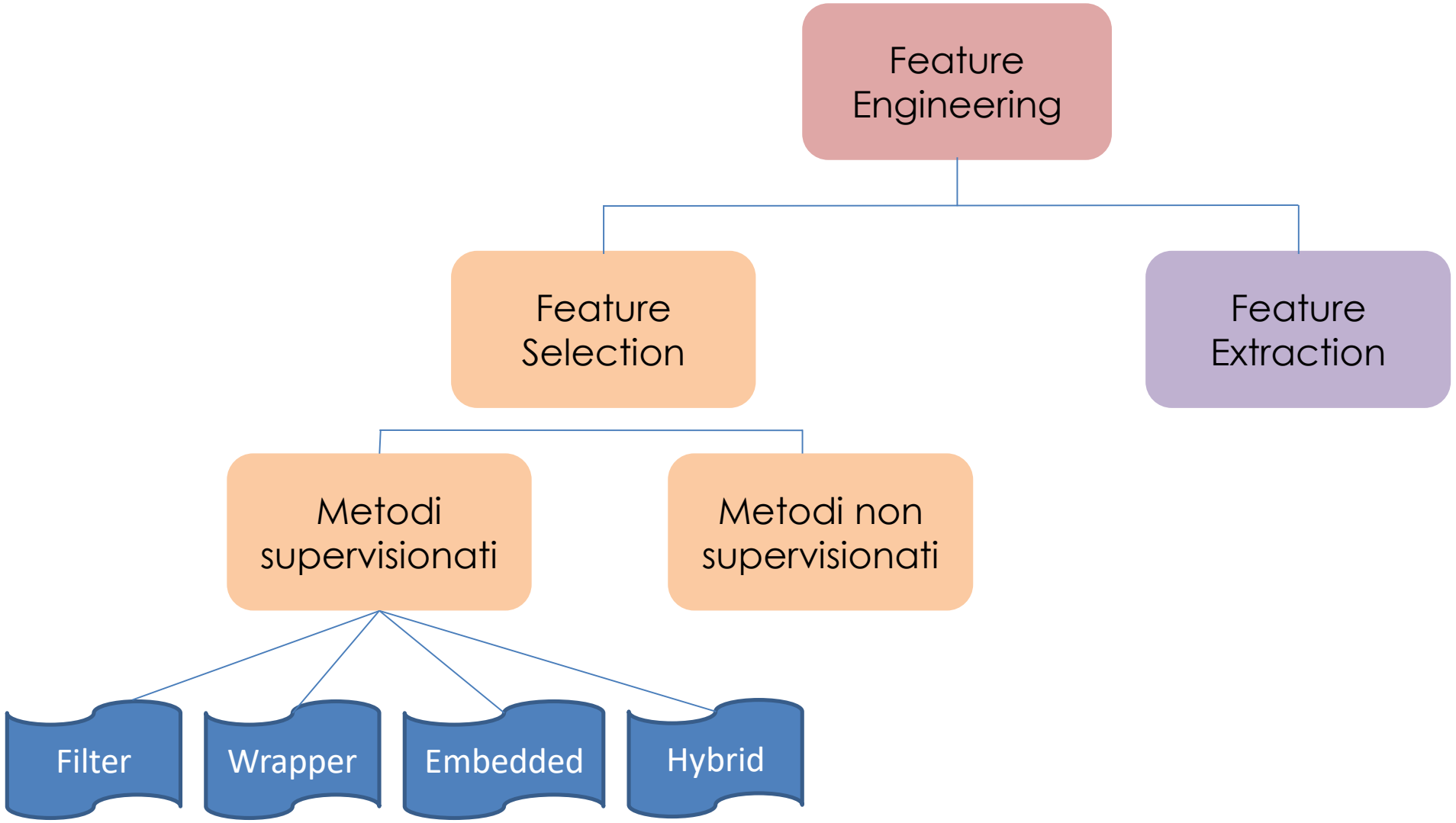
4. Riduce il tempo di training: meno variabili riducono la complessità dell'algoritmo, di conseguenza gli algoritmi si eseguono e si addestrano più velocemente.

5. Risolve il problema dello squilibrio di classe, quando una delle classi ha più campioni rispetto alle altre classi.

Com'è fatto un algoritmo di selezione delle feature?



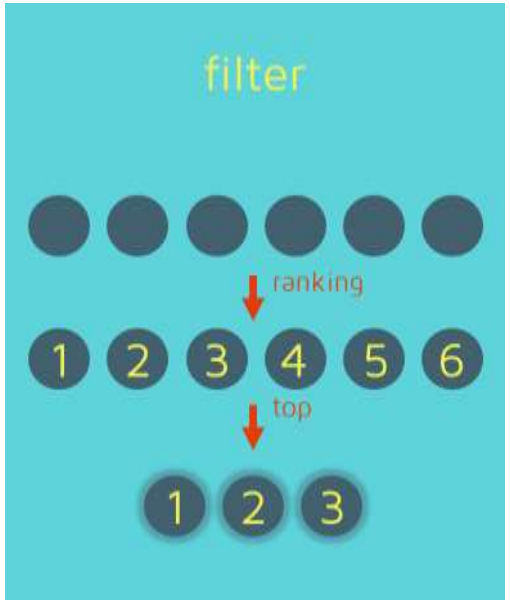
Feature selection - metodi



Filter methods

Consentono di filtrare le feature sulla base di una metrica statistica (es: Information Gain, Fisher's score, Chi quadro, ecc…).

In base alla metrica statistica selezionata ad ogni colonna di feature viene assegnato un punteggio che consente di individuare le feature rilevanti all'interno del dataset e di scartare quelle irrilevanti e ridondanti.



Vantaggi:

- Basso costo computazionale

- Buona capacità di generalizzazione

Filter methods

- Information Gain (IG)
- Fisher score
- Coefficiente di correlazione
- Chi-squared test (χ^2)

- Variance threshold
- Mean Absolute Deviation (MAD)
- Dispersion ratio

Filter method

INFORMATION GAIN (IG)

L'information gain (guadagno d'informazione) viene utilizzato per quantificare la riduzione dell'entropia che si verifica quando un dataset viene trasformato. Questo metodo confronta i valori di entropia prima e dopo della trasformazione.

L'IG applicata alla selezione delle variabili è chiamata **mutual information (MI)** (informazione reciproca) e quantifica la dipendenza statistica tra due variabili selezionate per far parte del nuovo sottoinsieme oppure può essere usato per valutare il guadagno di ogni variabile rispetto alla variabile target.

Filter methods – FISHER SCORE

Il Fisher score è uno dei metodi supervisionati di selezione delle feature più utilizzati, soprattutto nei problemi di classificazione binaria.

Questo algoritmo calcola un punteggio secondo il criterio di Fisher per ogni singola feature indipendentemente, dopodichè restituisce una classifica delle variabili in base al punteggio in ordine decrescente.

Limiti

Il principale limite di questo metodo è che non tiene conto della correlazione tra le variabili.

Filter methods – COEFFICIENTE DI CORRELAZIONE

La correlazione definisce la relazione reciproca tra due o più caratteristiche.

Come aiuta la correlazione nella selezione delle feature?

Le buone variabili dovrebbero essere altamente correlate con il target ma non dovrebbero essere correlate tra loro.

Se due variabili sono correlate, possiamo prevedere una dall'altra. Inoltre, il modello avrà bisogno solo di una di esse, poiché la seconda non aggiunge ulteriori informazioni.

Filter methods – COEFFICIENTE DI CORRELAZIONE

Correlazione di Pearson o p-value

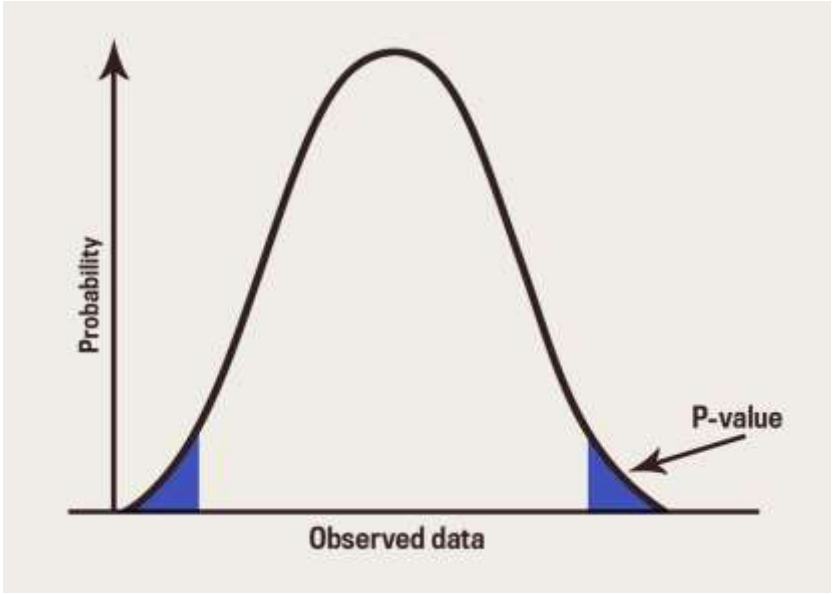
Ipotesi nulla: è un'affermazione generale che dice che non c'è relazione tra due fenomeni misurati.

Es. - H0 :- Non c'è relazione tra la caratteristica indipendente e la variabile obiettivo

- H1 :- C'è qualche relazione tra la caratteristica indipendente e la variabile obiettivo

Impostiamo 0.5 come valore soglia assoluto di selezione delle variabili

- Se p-value ≥ 0.5 non è possibile rifiutare l'ipotesi nulla
- Se p-value < 0.5 è possibile rifiutare l'ipotesi nulla



Filter method - CHI-SQUARED TEST (χ^2)

Il test del Chi-quadrato è usato per le features categoriche. Si calcola il χ^2 tra ogni feature e la variabile target e si seleziona il numero desiderato di features con i migliori punteggi di χ^2.

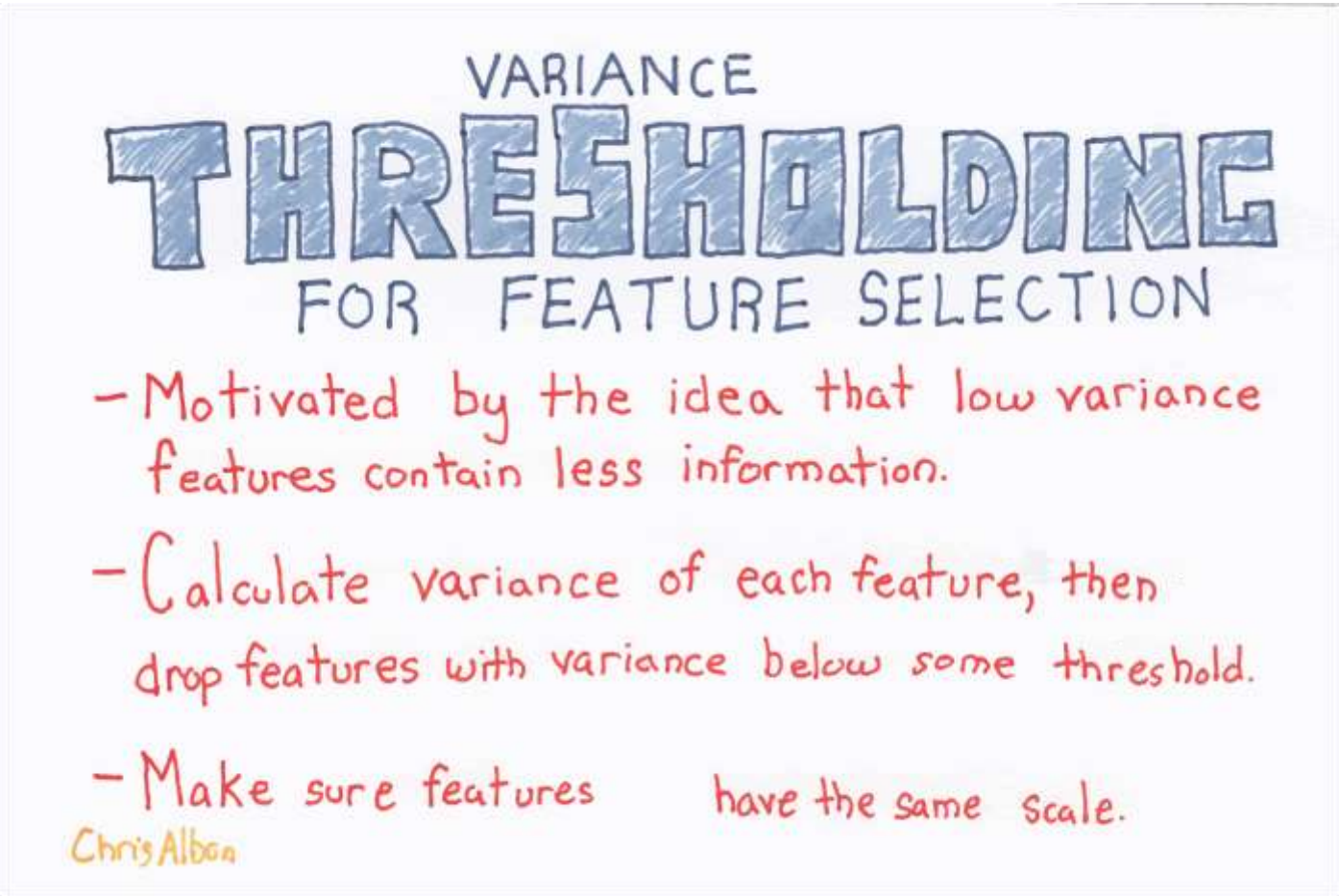
Nella selezione delle feature, miriamo a selezionare quelle che sono altamente dipendenti dalla risposta.

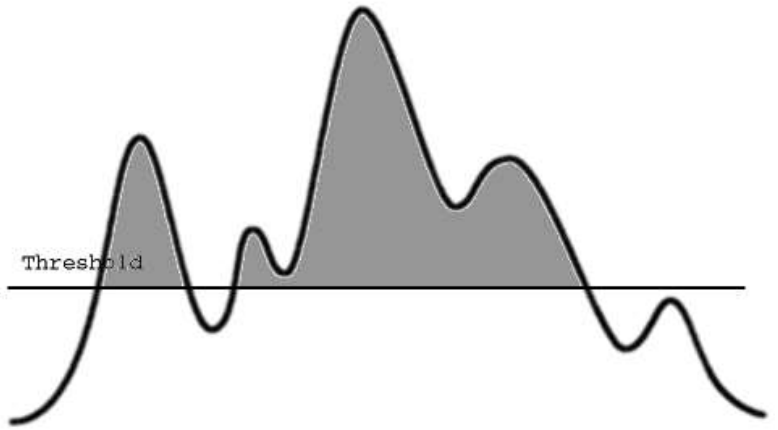
- χ^2 piccolo: feature indipendenti
- χ^2 elevato: feature dipendenti

Limiti

- ❑ le variabili devono essere categoriche
- ❑ campionate indipendentemente
- ❑ i valori devono avere una frequenza attesa maggiore di 5.

Filter methods – VARIANCE THRESHOLD





Limiti: non considera la correlazione tra variabili indipendnenti né la correlazione tra variabili indipendnenti e variabile target.

Filter methods - MEAN ABSOLUTE DEVIATION (MAD)

La deviazione assoluta media (MAD) è la media delle deviazioni assolute dei dati dalla media dei dati: la distanza media (assoluta) dalla media.

La MAD è una misura di variabilità più semplice della deviazione standard ed è quindi più facile da capire. MAD è più comunemente usato perché è più facile da calcolare (non necessita di elevazione al quadrato).

Filter methods – DISPERSION RATIO

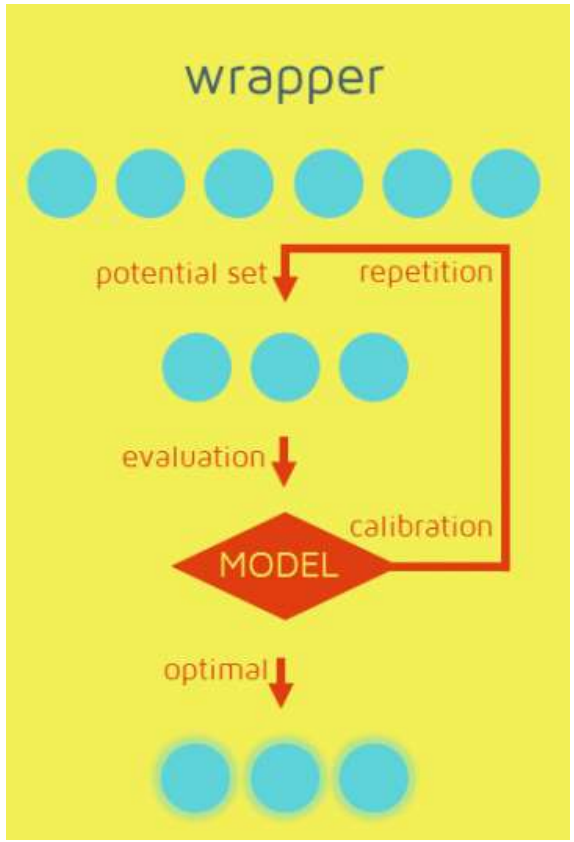
L'indice di dispersione è definito come il rapporto tra la media aritmetica (AM) e la media geometrica (GM) per una data feature. Il suo valore varia da +1 a ∞ con AM ≥ GM.

Un rapporto di dispersione più alto implica una caratteristica più rilevante.

Wrapper methods

La metodologia Wrapper considera la selezione delle feature come un problema di ricerca. Vengono preparati diversi sottoinsiemi di feature, valutati (rispetto ad un criterio di valutazione) attraverso un classificatore e confrontati fra loro.

Il miglior sottoinsieme di feature viene selezionato in base ai risultati del classificatore.



Vantaggi: migliore accuratezza predittiva rispetto ai metodi filter

Svantaggi: rischio di adattamento eccessivo quando le osservazioni sono insufficienti e il tempo di calcolo significativo quando il numero di variabili è elevato.

Wrapper methods

- Forward Feature Selection
- Backward Feature Elimination

} Sequential Feature Selection

- Exhaustive Feature Selection
- Recursive feature elimination
- Algoritmi genetici

Wrapper methods – FORWARD FEATURE SELECTION

Metodo iterativo che inizia con nessuna feature nel modello, la prima feature che viene aggiunta è la variabile più performante rispetto alla variabile target. Successivamente, l'algoritmo seleziona un'altra variabile che dà le migliori prestazioni in combinazione con la prima variabile selezionata. Questo processo continua fino a che l'aggiunta di una variabile non fornisce alcun miglioramento per la performance del modello.

Wrapper methods – BACKWARD FEATURE ELIMINATION

Questo metodo funziona esattamente all'opposto del metodo Forward Feature Selection.

Qui, iniziamo con un insieme che comprende tutte le feature disponibili e costruiamo un modello. Successivamente, ad ogni iterazione si elimina dal modello la variabile meno significativa la cui eliminazione migliora le prestazioni del modello.

Il processo continua finchè l'eliminazione dell'ultima feature non produce alcun miglioramento nelle prestazioni del modello.

Wrapper methods – EXAUSTIVE FEATURE SELECTION

Implica una valutazione a forza bruta di ogni sottoinsieme di feature.
Prova ogni possibile combinazione delle variabili e restituisce il sottoinsieme più performante.

Limite: elevati tempi di calcolo quando il dataset contiene moltissime feature

Wrapper methods – RECURSIVE FEATURE ELIMINATION (RFE)

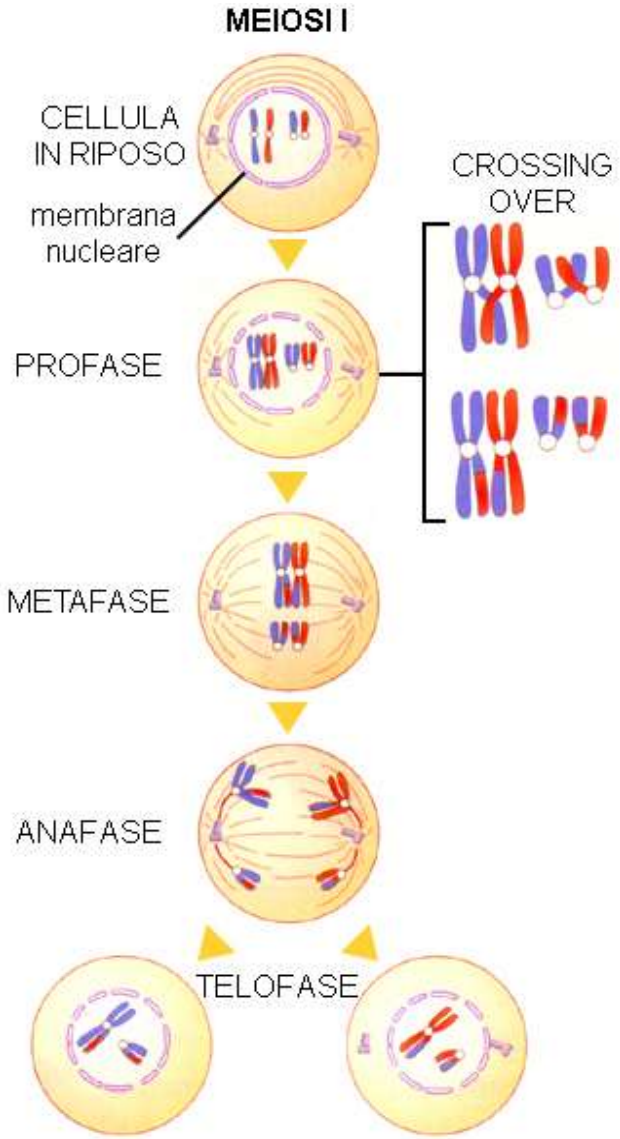
Questa tecnica assomiglia alla backward elimination, inizia costruendo un modello sull'intero set di feature e calcolando un punteggio di importanza per ogni feature.

A differenza della backward elimination però l'obiettivo RFE è di selezionare le feature considerando ricorsivamente insiemi sempre più piccoli di feature. Questa procedura viene ripetuta ricorsivamente sull'insieme potato fino a raggiungere il numero desiderato di feature da selezionare.

Algoritmi genetici

L'algoritmo genetico è un metodo stocastico ispirato al funzionamento della genetica e dell'evoluzione biologica.

In natura, i nostri geni tendono ad evolversi nelle generazioni successive per adattarsi meglio all'ambiente.



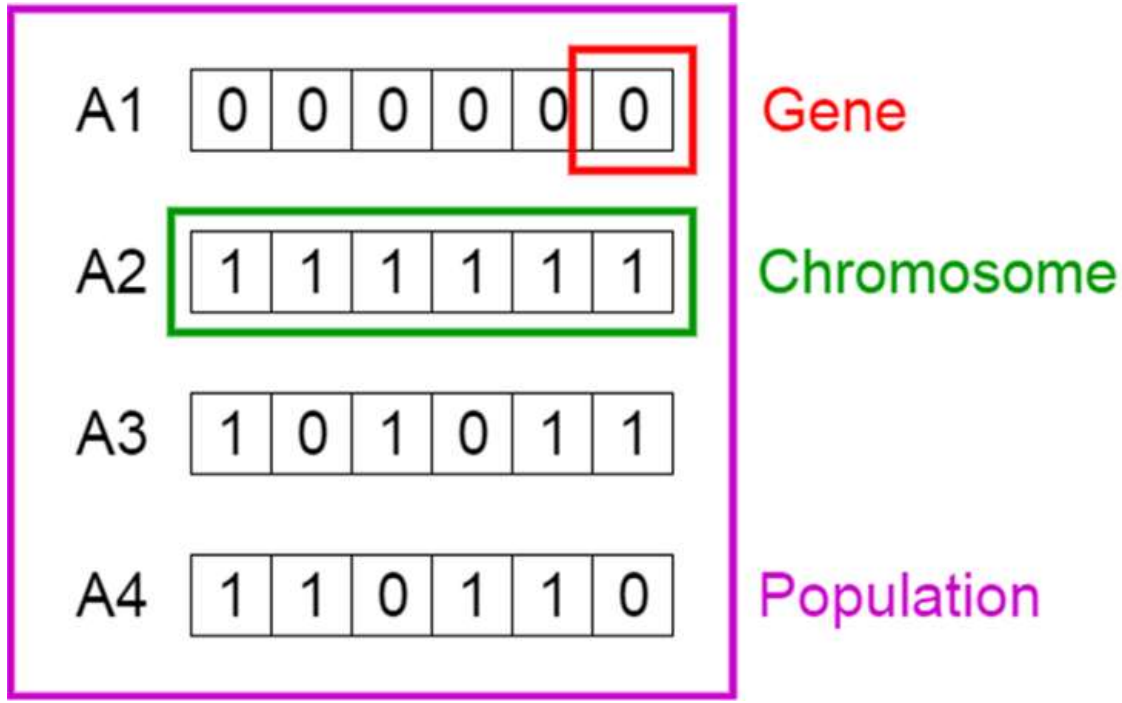
Algoritmi genetici

Nella feature selection

Popolazione: tutte le possibili combinazioni di feature del dataset

Cromosoma: una specifica combinazione di geni (feature), e la combinazione è rappresentata da una stringa che ha una lunghezza pari al numero totale di feature.

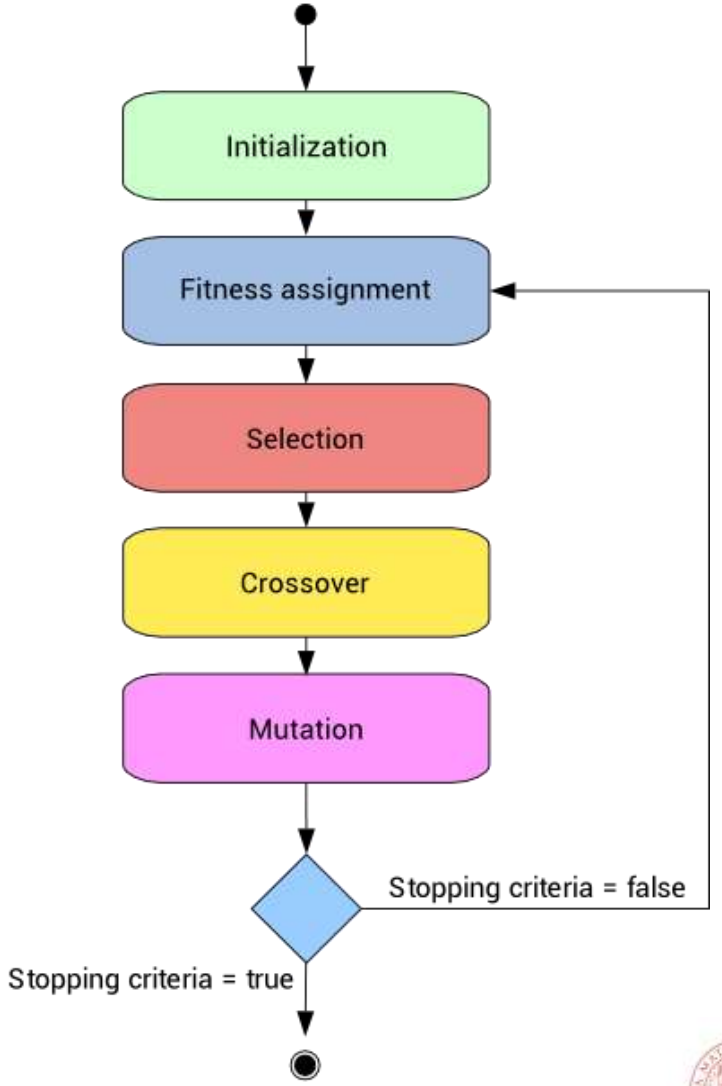
Gene: ogni singola feature



Algoritmi genetici

1. Iniziazione Ogni singolo membro della popolazione (il sottoinsieme di feature) rappresenta la soluzione al problema che deve essere risolto. Ogni individuo (sottoinsieme) è associato a parametri noti come geni (features) che sono uniti in una stringa per formare un cromosoma.

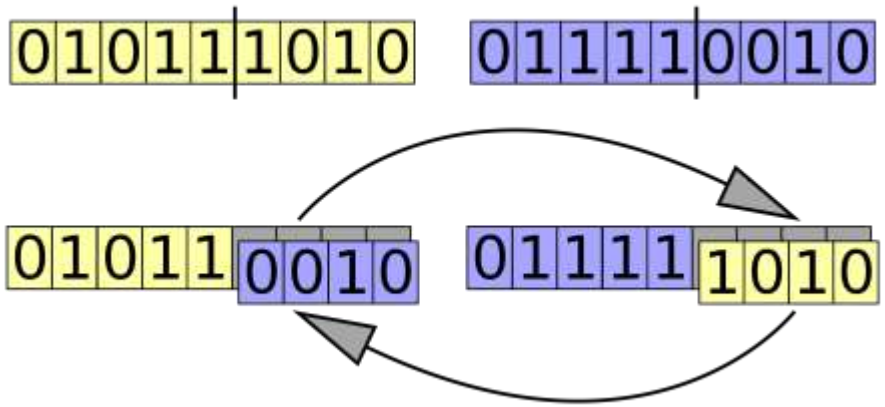
2. Fitness assignement Una volta che la generazione iniziale è stata creata, viene stimato il fitness (o capacità predittiva) di ogni sottoinsieme di caratteristiche. Il **criterio di fitness** è la capacità predittiva della combinazione di feature selezionata. I sottoinsiemi di caratteristiche nei genetic algorithm sono raggruppati in generazioni invece di considerare un sottoinsieme alla volta.

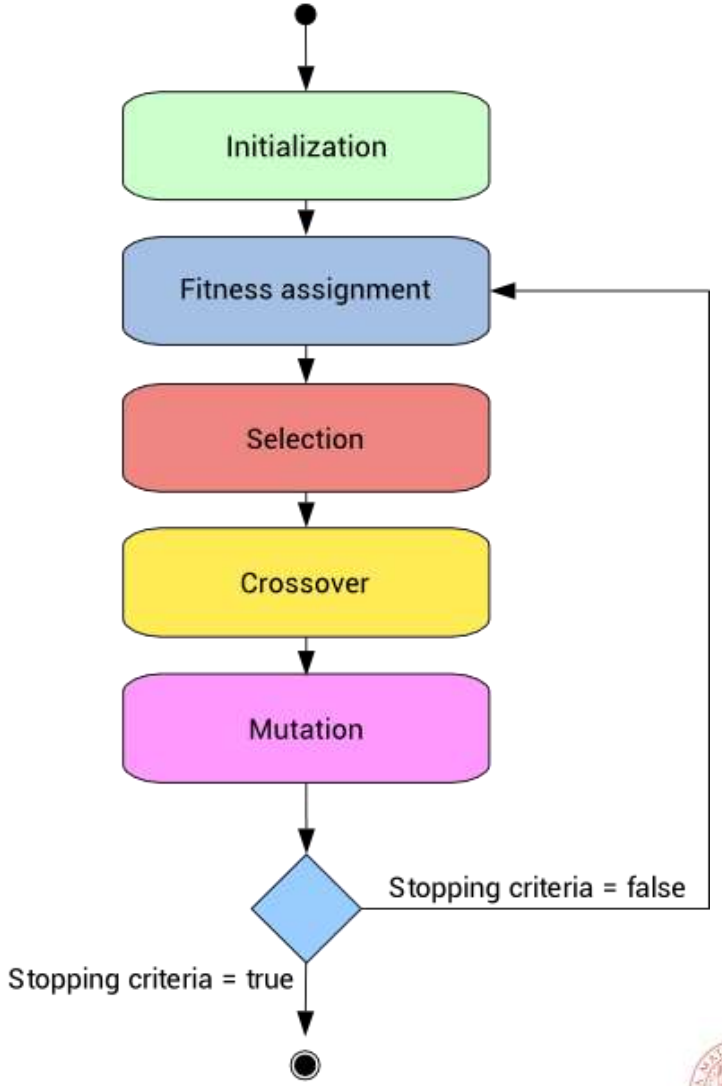


Algoritmi genetici

3. Selezione: in questa fase si decide quale membro della generazione esistente viene selezionato per produrre la nuova generazione e questo si basa sulla funzione fitness di quell'individuo.

4. Crossover (ricombinazione): è un'operazione di fusione delle informazioni genetiche dei genitori (padre e madre) per produrre nuova prole. Nel crossover a punto singolo viene scelta una posizione casuale tra due cromosomi. I genitori si scambiano le caratteristiche per generare due nuovi individui.





Algoritmi genetici

5. Mutazione : la disposizione dei geni nel cromosoma si altera per produrre un cromosoma totalmente nuovo.

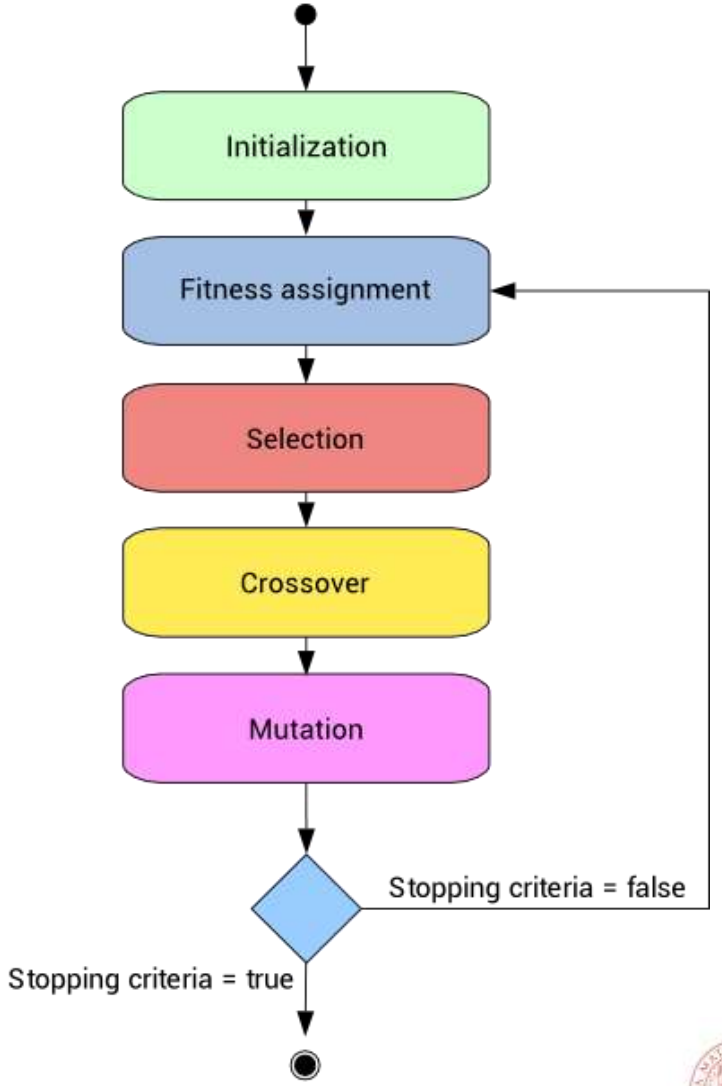
Before Mutation

A5	1	1	1	0	0	0

After Mutation

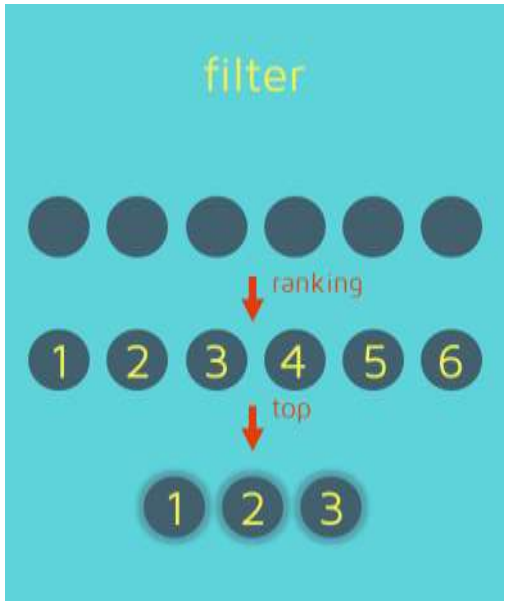
A5	1	1	0	1	1	0

6. Terminazione : Come sapete, un algoritmo termina ad un certo punto. Questo processo è anche chiamato convergenza. L'algoritmo genetico si ripete fino a quando non si raggiungono certe condizioni in cui la nuova generazione non è molto diversa dalla precedente. Quando non si verificano molti nuovi cambiamenti l'algoritmo termina.

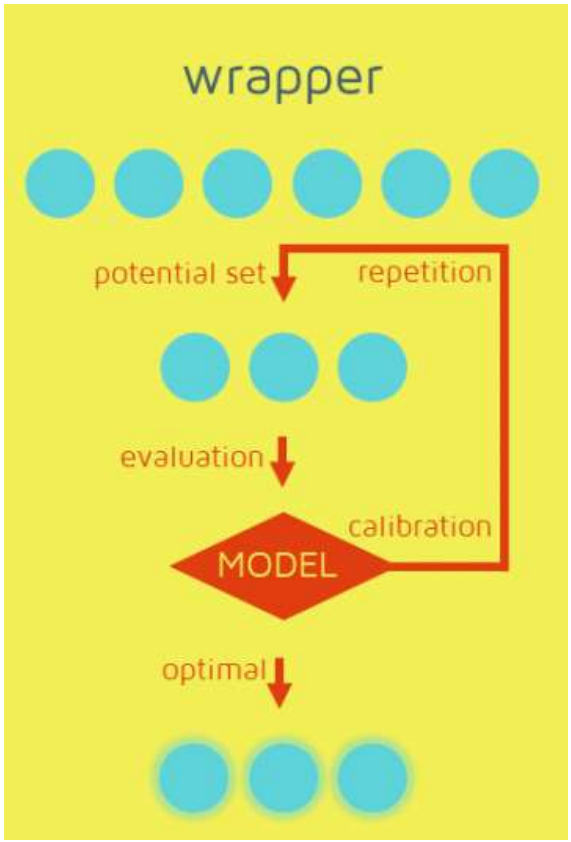


Hybrid methods

I metodi ibridi combinano i vantaggi dei metodi filtro e dei wrapper



+

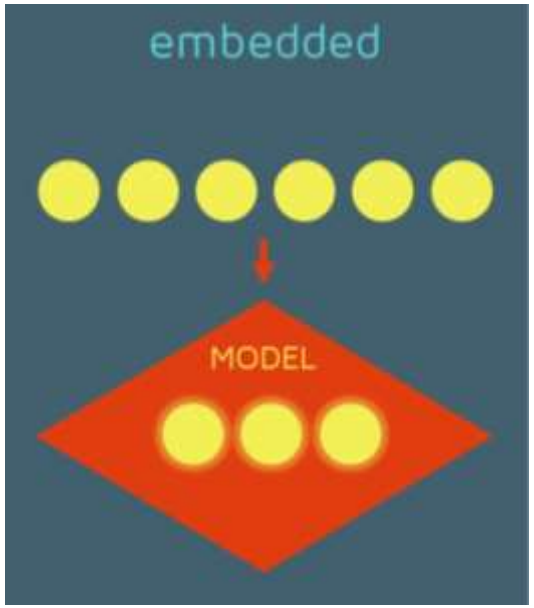


Embedded methods

L'algoritmo di selezione delle feature è integrato come parte dell'algoritmo di apprendimento (es: decision trees, random forest).

I metodi embedded presentano le qualità dei metodi filtro e wrapper.

Un algoritmo di apprendimento sfrutta il proprio processo di selezione delle variabili ed esegue la selezione delle feature e la classificazione/regressione allo stesso tempo.



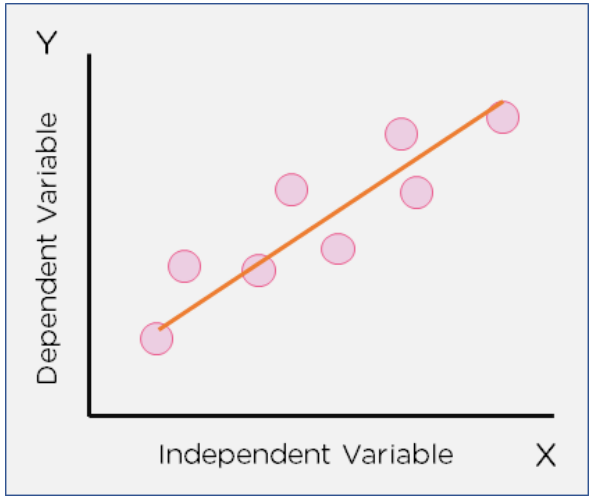
Vantaggi:

- Prendono in considerazione l'interazione delle caratteristiche come fanno i metodi wrapper.
- Sono più veloci come i metodi filter, ma più accurati.
- Sono molto meno inclini all'over-fitting.

Embedded methods

- Lasso regularization (L1)
- Random forest importance

Embedded methods - LASSO REGULARIZATION (L1)



La regressione Lasso è un algoritmo di regressione lineare che utilizza il restringimento (shrinkage), infatti l'acronimo "LASSO" sta per Least Absolute Shrinkage and Selection Operator.

La regressione Lasso esegue la regolarizzazione L1 la quale permette di fare feature selection.

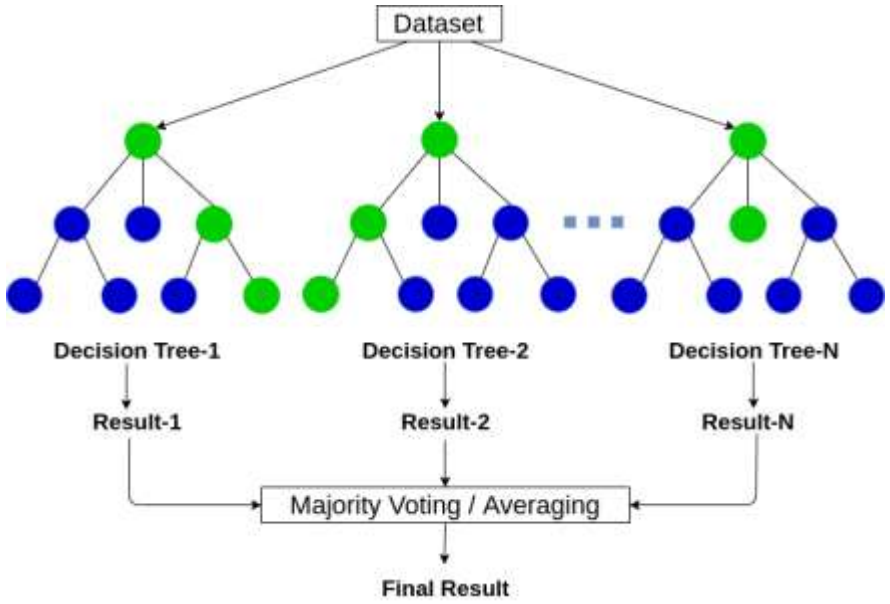
La regolarizzazione consiste nell'aggiungere una penalità ai parametri sui coefficienti che moltiplicano ciascuno dei predittori del modello di machine learning per ridurrne la libertà (evitare l'overfitting).

Lasso o L1 è in grado di ridurre alcuni dei coefficienti a zero. Pertanto, quella caratteristica può essere rimossa dal dataset.

Embedded methods - RANDOM FOREST IMPORTANCE

Random Forest è un algoritmo che aggrega un numero specificato di alberi di decisione.

Le Random Forest sono composte alberi decisionali, ognuno dei quali è costruito su un'estrazione random (casuale) di samples e feature. Ogni albero è una sequenza di domande sì-no basate su una singola o una combinazione di caratteristiche. Ad ogni nodo (cioè ad ogni domanda), gli alberi dividono il dataset in 2 nodi, ognuno dei quali ospita osservazioni che sono più simili tra loro e diverse da quelle dell'altro nodo.



Pertanto, l'importanza di ogni caratteristica deriva da quanto è "puro" ciascuno dei nodi.

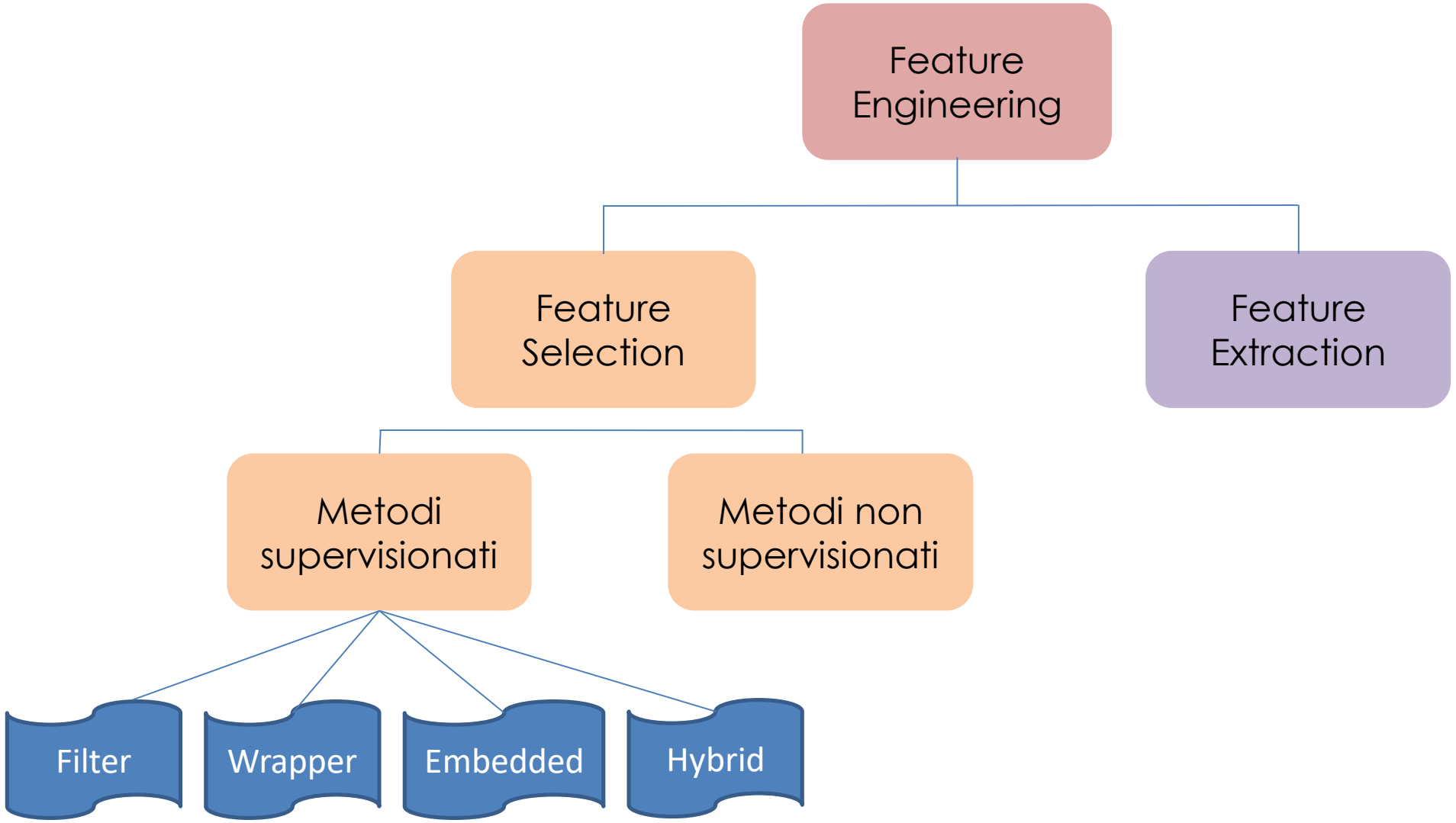
Come si misura la purezza?

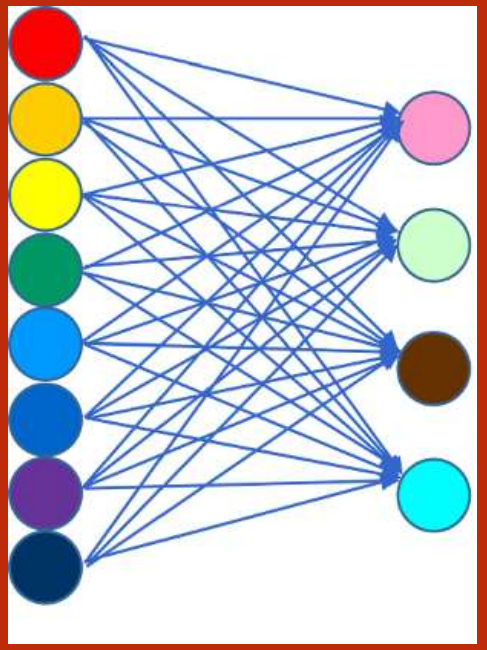
Classificazione: impurità di Gini o il information gain.

Regressione: varianza.

Più una caratteristica diminuisce l'impurità, più importante è la caratteristica.

Feature selection - metodi





Feature Extraction

Valentina Pellicioni

Studente di Dottorato in «Scienza e Cultura del Benessere e degli Stili di Vita»

Dipartimento di Scienze per la Qualità della vita

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Feature extraction

La Feature Extraction è un processo di riduzione della dimensionalità con cui un set iniziale di dati viene ridotto in un gruppo più gestibile per l'elaborazione.

Perché è utile?

L'idea alla base è quella di comprimere i dati dallo spazio iniziale ad uno spazio con dimensione inferiore mantenendo la maggior parte delle informazioni rilevanti.

Scopo:

ridurre il numero di variabili dal set originale per ridurre la complessità del modello, l'overfitting, migliorare l'efficienza del calcolo e ridurre l'errore di generalizzazione.

Tecniche di feature extraction

- Principal Component Analysis (PCA)
- Linear Discriminant Analysis (LDA)

PCA e LDA sono due popolari metodi di riduzione della dimensionalità comunemente usati su dati ad alta dimensionalità. Per molti versi i due algoritmi sono simili, ma allo stesso tempo molto dissimili.

La **PCA** è una tecnica di trasformazione lineare **non supervisionata**, riduce il numero di dimensioni trovando la massima varianza nei dati.

L'**LDA** è un metodo **supervisionato** che prende in considerazione le etichette di classe quando riduce il numero di dimensioni. L'obiettivo di LDA è di trovare un sottospazio di caratteristiche che ottimizzi al meglio la separabilità delle classi.

LDA e PCA a confronto

LDA

1 - Per ogni etichetta di classe calcolare il vettore medio d-dimensionale

2 - Costruire una matrice di dispersione all'interno di ogni classe e tra ogni classe.

3 - Calcola la miglior proiezione LDA: Per trovare la migliore proiezione, dobbiamo calcolare
gli *autovettori* e *autovalori* corrispondenti.

4 - Sceglie i migliori autovalori: Una volta trovati gli autovalori, li ordiniamo in modo decrescente e selezioniamo il superiore

5 - Crea una matrice di autovettori trovati: Creiamo una nuova matrice contenente autovettori che mappano gli autovalori

6 - Ottieni le nuove feature dalla LDA: ovvero le nuove componenti

PCA

1 - Costruire la matrice di covarianza prendendo la covarianza congiunta tra ogni coppia nel vettore dato.

Poi con la matrice generata...

1 - Calcolare gli autovettori e gli autovalori della matrice

2 - Ordinare gli autovalori per ordine decrescente per classificare gli autovettori

3 - Ottenere i k autovettori che corrispondono ai k maggiori autovalori

4 - Costruire una matrice di proiezione dai primi k autovettori

5 - Trasformare il set di dati di input originale con la proiezione appena creata

Passi cruciali dell'algoritmo PCA

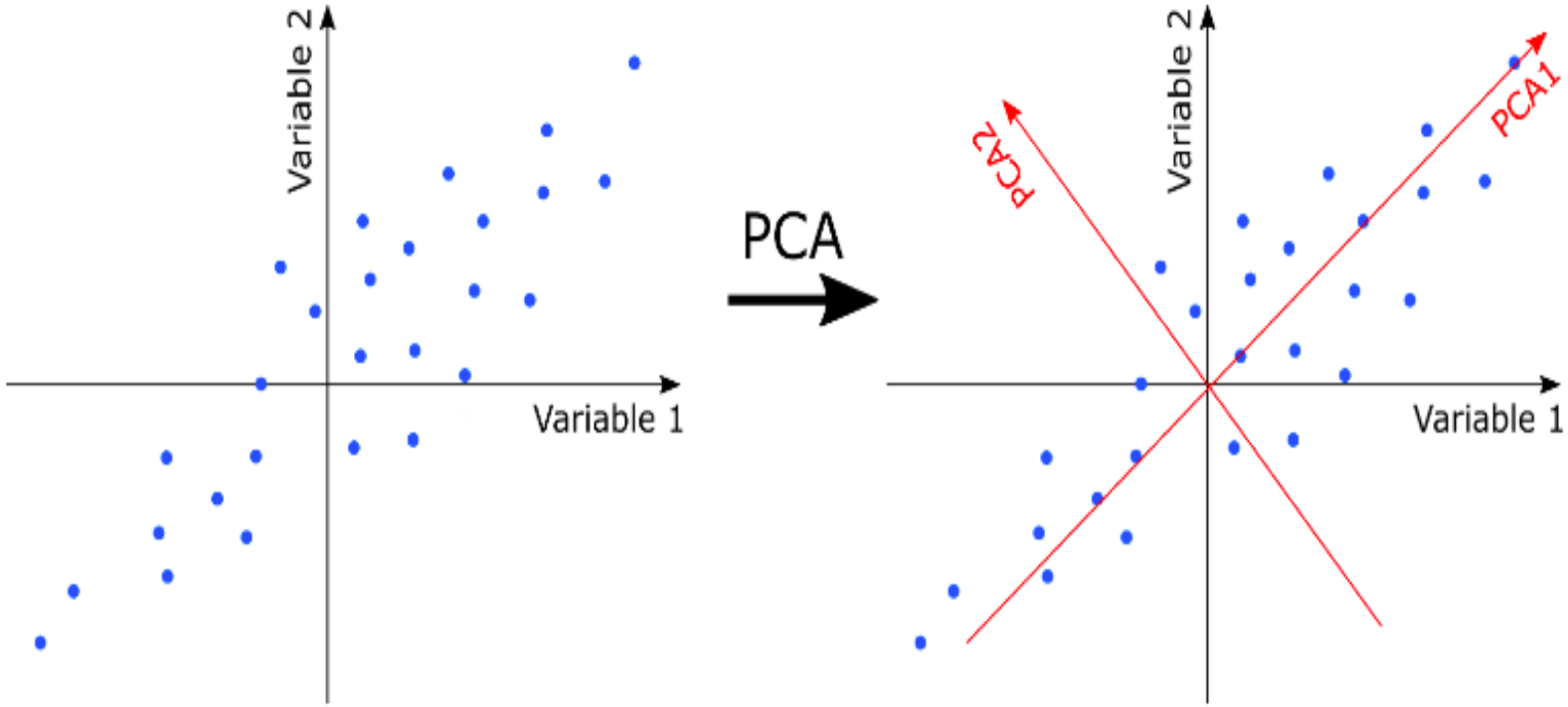
1. Standardizzare i dati: è il processo di ridimensionamento di uno o più attributi in modo che abbiano un valore medio di 0 e una deviazione standard di 1.

2. Calcolare la matrice di covarianza per identificare le correlazioni: capire come le variabili del dataset di input variano dalla media rispetto alle altre, o in altre parole, vedere se c'è qualche relazione tra loro.

3. Calcolare gli autovettori e gli autovalori della matrice di covarianza per identificare le componenti principali: gli autovettori della matrice di covarianza sono le direzioni degli assi dove c'è più varianza (più informazione) e che chiamiamo Componenti Principali. Gli autovalori sono i coefficienti legati agli autovettori, che indicano la quantità di varianza trasportata in ogni componente principale.

4. Creare un vettore di feature per decidere quali componenti principali mantenere: In questo passo, scegliamo se mantenere tutte le componenti o scartare quelle di minore importanza, e formare con quelle rimanenti una matrice di vettori che chiamiamo Feature vector.

5. Formare le componenti principali



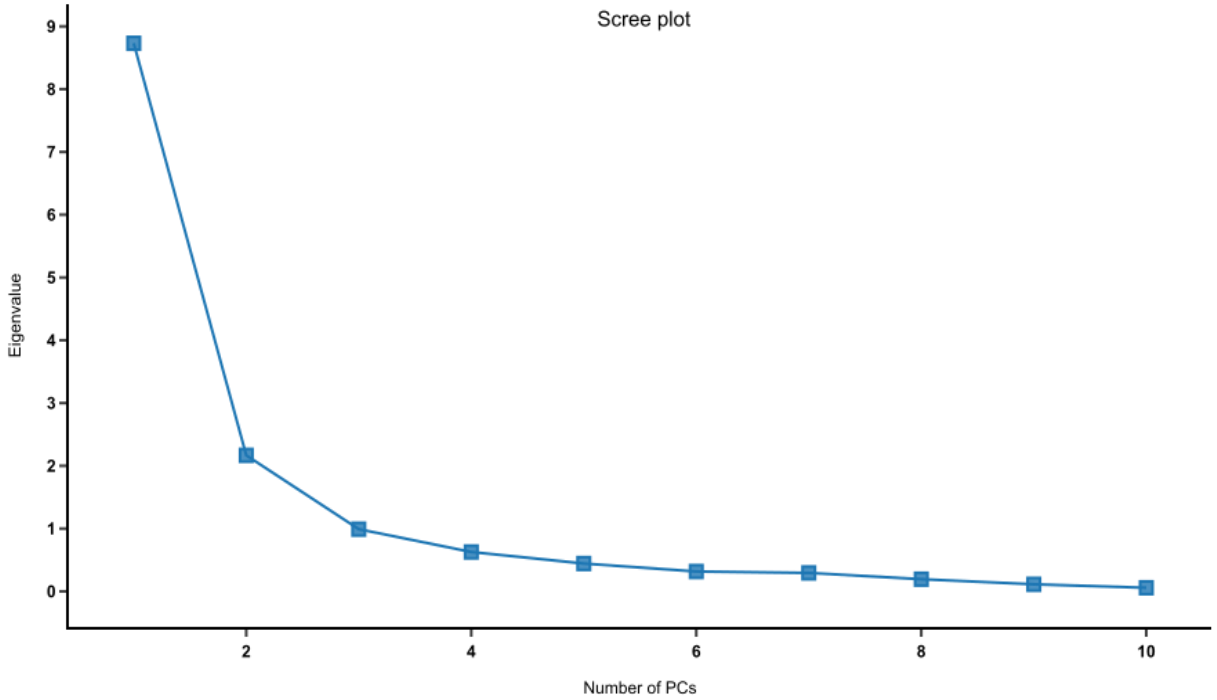
Come si determina il numero di componenti principali (PC)?

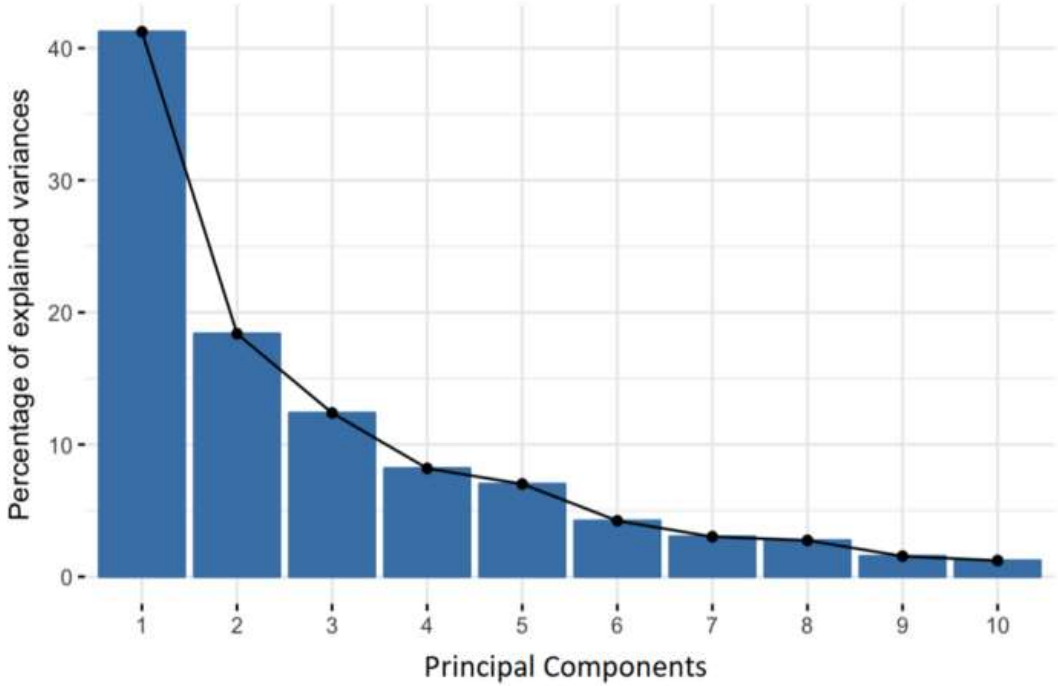
- Criterio dell'autovalore
- Criterio della percentuale di varianza spiegata
- Criterio dello Scree plot

Come si determina il numero di componenti principali (PC)?

SCREE PLOT

Scree plot





Elbow rule: in uno scree plot ideale la discesa della curva dovrebbe essere ripida, poi piegarsi a "gomito" e infine appiattirsi. Il punto in cui si forma il gomito (elbow) indica il numero di PC che dovremmo considerare.

Altre tecniche di feature extraction

- Independent Component Analysis (ICA)
- Locally Linear Embedding (LLE)
- t-distributed Stochastic Neighbor Embedding (t-SNE)
- Autoencoders

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Valentina Pellicioni

Dipartimento di Scienze per la Qualità della Vita

valentina.pellicion2@unibo.it