

Machine Learning: Algoritmi e Modelli

Nearest Neighbors

Introduzione ai Nearest Neighbors

presentato da
Giuseppe Gullo

PROFESSION 

Algoritmi Nearest Neighbors

Sono una famiglia di algoritmi che porta alla realizzazione di **modelli lazy**,
cioè modelli che non apprendono dai dati ma li memorizzano

Algoritmi Nearest Neighbors

Sono una famiglia di algoritmi che porta alla realizzazione di **modelli lazy**,
cioè modelli che non apprendono dai dati ma li memorizzano

Quello che fanno è trovare un **numero predefinito** di osservazioni
più vicini al nuovo punto e prevedere il label da questi.

Algoritmi Nearest Neighbors

Sono una famiglia di algoritmi che porta alla realizzazione di **modelli lazy**,
cioè modelli che non apprendono dai dati ma li memorizzano

Quello che fanno è trovare un **numero predefinito** di osservazioni
più vicini al nuovo punto e prevedere il label da questi.



Proximity Search

Algoritmi Nearest Neighbors

Proximity Search

Dato un insieme S di punti in uno spazio M
e un punto di interrogazione $q \in M$,
trovare il punto di S più vicino a q

Algoritmi Nearest Neighbors

Proximity Search

Dato un insieme S di punti in uno spazio M
e un punto di interrogazione $q \in M$,
trovare il punto di S più vicino a q

Post-office problem



Machine Learning: Algoritmi e Modelli

Nearest Neighbors

KNN: K-Nearest Neighbors

presentato da
Giuseppe Gullo

PROFESSION 

KNN: K-Nearest Neighbors

E' l'algoritmo Nearest Neighbors più popolare.

La classe predetta è quella delle k osservazioni più vicine.

KNN: K-Nearest Neighbors

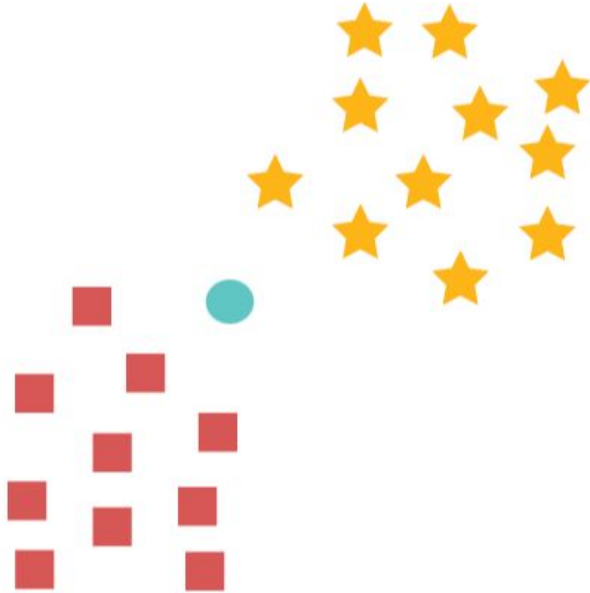
E' l'algoritmo Nearest Neighbors più popolare.

La classe predetta è quella delle k osservazioni più vicine.

K è un iperparametro:

- Valori maggiori riducono l'effetto del rumore
- Valori minori portano a decision boundaries più netti

Un esempio di KNN



Un esempio di KNN

$K = 5$



Passaggi

- Scegliamo il valore di K

Un esempio di KNN

$K = 5$

Distanza Euclidea



Passaggi

- Scegliamo il valore di K
- Scegliamo una metrica per la distanza

Un esempio di KNN

$K = 5$

Distanza Euclidea



Passaggi

- Scegliamo il valore di K
- Scegliamo una metrica per la distanza
- Cerchiamo i K vicini più prossimi all'osservazione che vogliamo classificare

Un esempio di KNN

$K = 5$

Distanza Euclidea



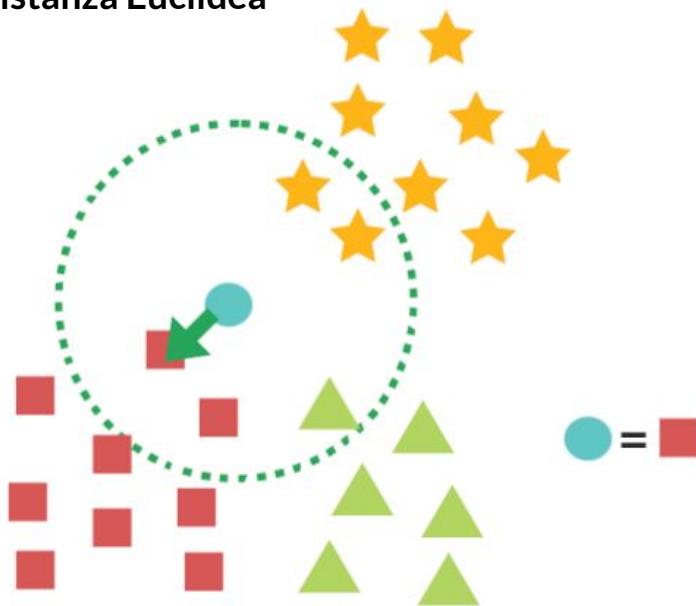
Passaggi

- Scegliamo il valore di K
- Scegliamo una metrica per la distanza
- Cerchiamo i K vicini più prossimi all'osservazione che vogliamo classificare
- Assegnamo la classe più presente tra i nearest neighbors

Un esempio di KNN

$K = 5$

Distanza Euclidea



Passaggi

- Scegliamo il valore di K
- Scegliamo una metrica per la distanza
- Cerchiamo i K vicini più prossimi all'osservazione che vogliamo classificare
- Assegniamo la classe più presente tra i nearest neighbors
- Se c'è un pareggio tra le classi assegniamo la classe con la distanza minore

Pro del K-Nearest Neighbors

- Estremamente semplice
- Si adatta subito a nuovi dati
- Può apprendere decision boundary molto irregolari

Pro del K-Nearest Neighbors

- Estremamente semplice
- Si adatta subito a nuovi dati
- Può apprendere decision boundary molto irregolari



NON PARAMETRICO

Contro del K-Nearest Neighbors

- Dispendioso per dataset grandi
- Molto sensibile agli outlier
- Non interpretabile

Machine Learning: Algoritmi e Modelli

Nearest Neighbors

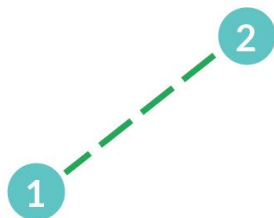
Metriche per la distanza

presentato da
Giuseppe Gullo

Distanza Euclidea

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

y



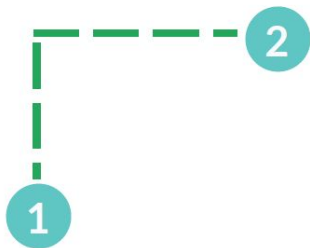
x

E' maggiormente influenzata dalla dimensione con la differenza più grande

Distanza di Manhattan

$$|x_1 - x_2| + |y_1 - y_2|$$

y



x

E' ugualmente influenzata da tutte le dimensioni

Distanza di Minkowski

$$\left(|x_1 - x_2|^p + |y_1 - y_2|^p \right)^{\frac{1}{p}}$$

Tramite p possiamo controllare il peso da dare alle differenze maggiori

Machine Learning: Algoritmi e Modelli

Nearest Neighbors

Radius Nearest Neighbors

presentato da
Giuseppe Gullo

PROFESSION 

RNN: Radius Nearest Neighbors

Algoritmo Nearest Neighbors utilizzato quando le osservazioni sono distribuite in maniera non uniforme.

Si concentra su una regione di raggio R e non su un numero di osservazioni.

La classe predetta è quella del maggior numero di osservazioni contenute in un raggio R .

RNN: Radius Nearest Neighbors

Algoritmo Nearest Neighbors utilizzato quando le osservazioni sono distribuite in maniera non uniforme.

Si concentra su una regione di raggio R e non su un numero di osservazioni.

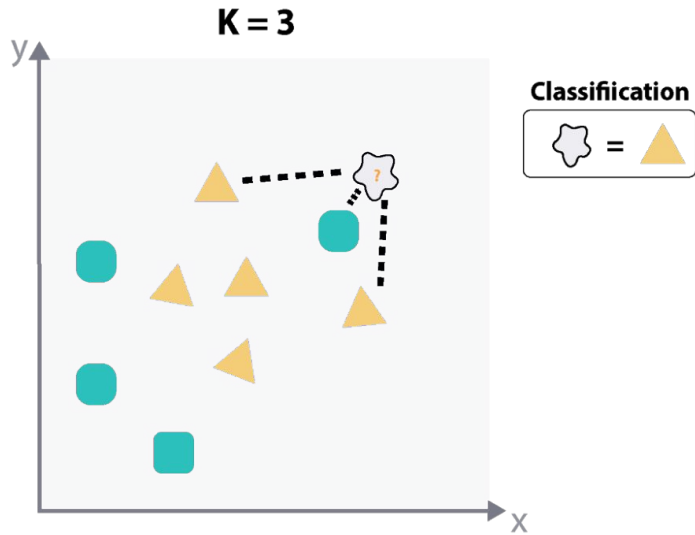
La classe predetta è quella del maggior numero di osservazioni contenute in un raggio R .

R è un iperparametro:

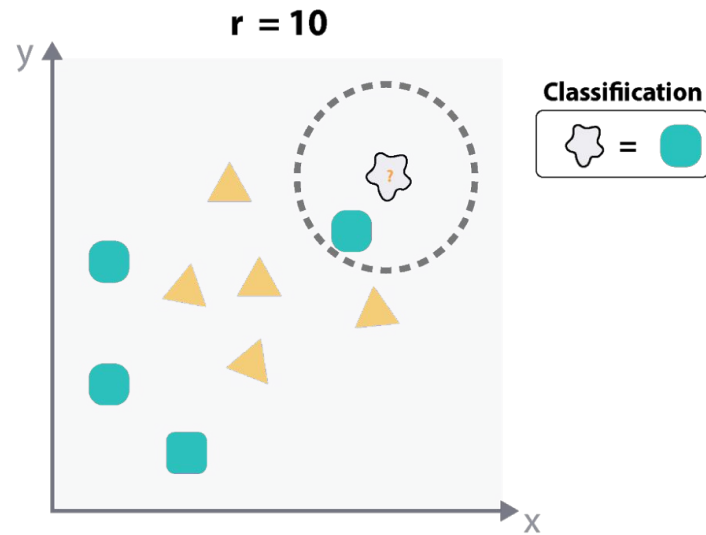
- Valori maggiori riducono l'effetto del rumore
- Valori minori portano a decision boundaries più netti

KNN VS RNN

KNN



RNN



KNN VS RNN

- RNN è più resistente al rumore
- KNN porta ad una stima più precisa se si sceglie un buon K