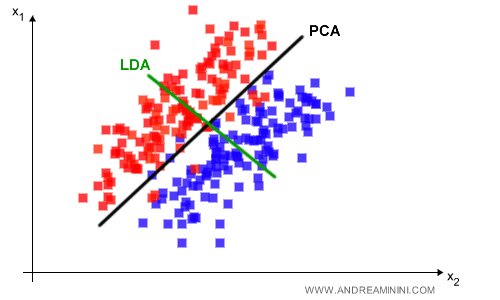
<https://www.andreaminini.com/ai/machine-learning/riduzione-dimensionalita-dati#:~:text=Principal%20Component%20Analysis%20(PCA),che%20rappresentano%20meglio%20i%20pattern>.

La riduzione della dimensionalità ( dimensionality reduction ) è una tecnica di mapping dei dati. E' una delle operazioni di pre-elaborazione del dataset nell'apprendimento automatico non supervisionato. E’ utile nel machine learning per eliminare dal dataset le informazioni ridondanti ( correlate ), meno o poco rilevanti per il problema da risolvere. E' senza dubbio più semplice e meno dispendioso addestrare un algoritmo con uno spazio dati di dimensione inferiore. Quindi è una soluzione al problema della curse of dimensionality.  
  
La curse of dimensionality è un problema della dispersione dei pattern in un grande volume di dati. Nel machine learning si lavora con una grande mole di dati ( big data ). A causa dell'elevata dimensionalità, i pattern sono dispersi nel dataset come in un oceano di rumore e dati insignificanti. In queste circostanze diventa complesso trovare uno schema. L'algoritmo impiega più tempo (complessità temporale) e più memoria (complessità spaziale). La riduzione della dimensionalità riduce il volume dei dati in cui cercare, senza perdere le informazioni più rilevanti.  
  
Ridurre la dimensionalità dei dati non vuol dire soltanto eliminare alcune dimensioni (rumore) ma soprattutto combinare tra loro le informazioni ridondanti e correlate. Gli obiettivi dell'algoritmo sono i seguenti: Eliminare il rumore dei dati Combinare le informazioni correlate Un dataset viene di una dimensione iniziale Rn viene ridotto a uno spazio di dimensione inferiore Rk dove k  
  
Le principali tecniche di riduzione della dimensionalità dei dati sono le seguenti: Principal Component Analysis (PCA) La tecnica consiste in un mapping lineare dei dati non supervisionato. E' anche detta tecnica KL (Karhunen Loeve). L'obiettivo della tecnica è individuare le dimensioni che rappresentano meglio i pattern. Linear Discriminant Anaylis (LDA) La tecnica LDA è un mapping lineare dei dati supervisionato. L'obiettivo della tecnica LDA è individuare le dimensioni che discriminano meglio i pattern.



La riduzione della dimensionalità ha vantaggi e svantaggi.   
Vantaggi. La riduzione della dimensionalità comprime il volume dei dati, riducendo la complessità computazionale dell'algoritmo di apprendimento.  
Svantaggi. La riduzione della dimensionalità può degradare le informazioni e le prestazioni predittive dell'algoritmo di apprendimento.