

## MLlib como módulo de machine learning con pyspark

MLlib es una librería de aprendizaje de Spark que consiste en algoritmos de aprendizaje comunes y utilidades, incluyendo clasificación, regresión, clustering, filtrado colaborativo, reducción de dimensionalidad...

<http://spark.apache.org/docs/latest/mllib-guide.html>

Librería que viene con spark para temas de machine learning. Provee una serie de algoritmos de machine learning:

- Clasificación: logistic regression, linear support vector machines (SVM)
- Regresión: linear regression
- Clustering:k-means
- Collaborative filtering:alternating least squares (ALS)
- Optimización:stochastic gradient descent (SGD)

La biblioteca de aprendizaje escalable de Spark, que consta de algoritmos de aprendizaje y utilidades, incluyendo clasificación, regresión, clustering, filtrado colaborativo

- Modelos lineales de regresión (SVMs, regresión logística y lineal)
- Clasificadores de árbol de decisión
- Filtrado colaborativo con ALS
- K-Means agrupación,clustering
- SVD (descomposición de valores singulares) y PCA
- Descenso de gradiente estocástico




El paquete en el que podemos encontrar algunos de los algoritmos de machine learning es **spark.mllib**

El código fuente de este paquete se encuentra en el repositorio:

<https://apache.googlesource.com/spark/+/master/python/pyspark/mllib>

[apache](#) / [spark](#) / [master](#) / [.](#) / [python](#) / [pyspark](#) / **mllib**

tree: bd73c4c6361640da4eaf741a21f98b93b61a516d [[path history](#)] [[tgz](#)]

-  [\\_\\_init\\_\\_.py](#)
-  [classification.py](#)
-  [clustering.py](#)
-  [common.py](#)
-  [evaluation.py](#)
-  [feature.py](#)
-  [fpm.py](#)
-  [linalg/](#)
-  [random.py](#)
-  [recommendation.py](#)
-  [regression.py](#)
-  [stat/](#)
-  [tests.py](#)
-  [tree.py](#)
-  [util.py](#)