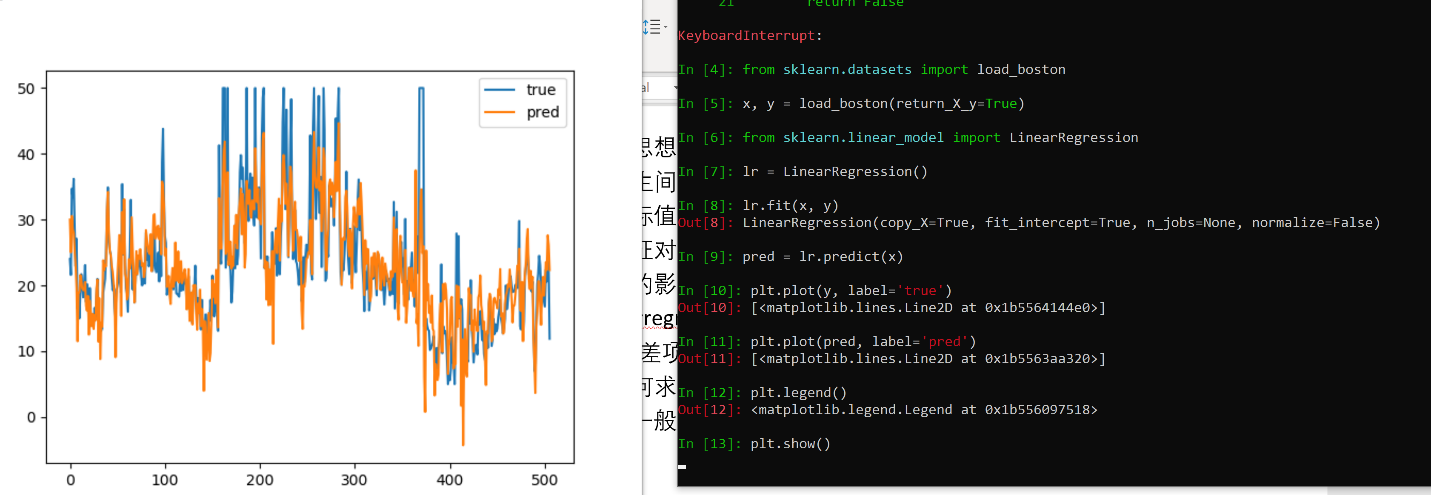
Common algorithms explain

1. Linear Regression:

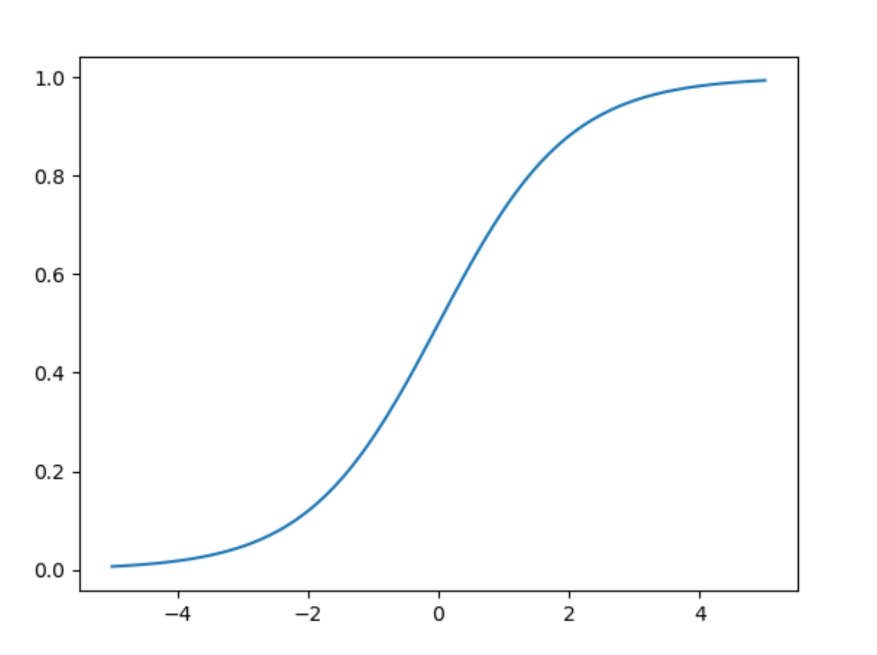
线性回归主要是一个回归算法，核心思想就是对特征的线性组合，假如我们想做房价预测，我们提取了3个特征如面积，卫生间的个数，卧室个数来描述一个房子的情况，我们的目标值数一个房子的价格，因为目标值是一个连续值，所以是一个回归问题，而我们可以知道对于不同的特征来说，每个特征对房价的影响是不一样的，比如面积影响比较大，卫生间的个数少，但对房价也有一定的影响，所以我们想通过一个模型来对过往房价的数据进行建模。所以我们可以使用linearregression来关联特征和目标值的关系。线性回归的核心就是w\*x + b， w为权重，b为偏差项，w对应不同的特征的重要程度，w比较大，则该特征对目标值影响比较大。而如何求w,b就是优化问题，通常我们使用梯队下降算法来对w,b进行求解，对于误差，我们一般使用RMSE均方误差评估。



1. Logistic Regression

逻辑回归是一个分类问题，本质上只能解决2分类问题，logisticregression本质上是建立在Linear Regression上，唯一的区别就是逻辑回归添加了一个sigmoid函数在线性回归的结果上，而sigmoid函数的输出结果在0-1 直接，可以代表概率，所以我们可以使用逻辑回归来进行分类。

Sigmoid函数



1. KNN

KNN 是一个分类算法，核心就是对于一个新样本的类别，我们可以计算新样本和全部的训练样本的距离来判断该新样本属于哪个类别，核心的参数k就是选择几个近邻，如果选择k=3，则新样本会计算和全部的样本的距离再排序，选择前3个最近的样本，然后计算3个样本分别的类别，找出那个类别出现最多的那个类别作为新样本的类别。

1. K-Means

Kmeans是一个聚类算法，聚类和分类的区别在于聚类没有样本的类别，所以是无监督学习。而kmeans核心是需要将现有数据聚成几类，如果k=2，则将全部数据聚成2类别；算法原理是随机从全部的样本中选择2个点作为2类的中心点，然后对全部的数据分别和2个随机选择的中心点求距离，对于每一个其他的点，都有2个距离分别和两个中心点的距离，选择距离最近的那个中心点作为该点的类别，遍历全部的样本点，则所有的样本点都有对应的类别，则将全部的数据分为2类，之后分别对2个类别分别求中心点，更新这个中心点为新的类别中心点，不断遍历，直到中心点不再变化或达到迭代次数。