|  |  |  |  |
| --- | --- | --- | --- |
| 類型 | 模型 | 演算法模型 | 分析法(問題意識) |
| Python Spark MLlib | 決策樹  二元分類 | DecisionTree.trainClassifier()  參數：   1. input=輸入的訓練資料 2. numClasses=2 3. categoricalFeaturesInfo=設定分類特徵欄位資訊 4. impurity=評估方法(gini or entropy) 5. maxDepth=最大深度 6. maxBins=每一節點的最大分支數目 | 判斷未知類別屬於哪種類别 |
| 邏輯回歸二元分類 | LogisticRegressionWithGSD.train()  參數：   1. input=輸入的訓練資料LabeledPointRDD 2. numIterations=SGD的迭代次數，預設為100 3. stepsize=執行迭代步伐大小，預設為1 4. miniBatchFraction=每次迭代參與計算的樣本比例，數值在0~1 |
| 支持向量機器SVM二元分類 | SVMWSGD.train()  參數：   1. input=輸入的訓練資料LabeledPointRDD 2. numIterations=SGD的迭代次數，預設為100 3. stepsize=執行迭代步伐大小，預設為1 4. miniBatchFraction=每次迭代參與計算的樣本比例，數值在0~1 |
| 單純貝氏二元分類 | NaiveBayesModel()  參數：   1. input=輸入的訓練資料LabeledPointRDD 2. lambda=設定lambda參數，預設值1.0 |
| Spark ML pipeline | 二元分類 | DecisionTreeClassifier()  參數：   1. labelCol=標籤欄位 2. featureCol=特徵欄位 3. impurity=評估方法(gini or entropy) 4. maxDepth=最大深度 5. maxBins=每一節點的最大分支數目 |
| Python Spark MLlib | 決策樹  多元分類 | DecisionTree.trainClassifier()  參數：   1. input=輸入的訓練資料 2. numClasses=any number >2 3. categoricalFeaturesInfo=設定分類特徵欄位資訊 4. impurity=評估方法(gini or entropy) 5. maxDepth=最大深度 6. maxBins=每一節點的最大分支數目 | 例：預測不同條件下，海拔、方位、斜率、與水源垂直距離、荒野分類等等，適合種植何種樹 |
| Spark ML pipeline | 多元分類 | DecisionTreeClassifier()  參數：   1. labelCol=標籤欄位 2. featureCol=特徵欄位 3. maxDepth=最大深度 4. maxBins=每一節點的最大分支數目 |
| Python Spark MLlib | 決策樹  回歸分析 | DecisionTree.trainRegressor()  參數：   1. input=輸入的訓練資料 2. categoricalFeaturesInfo=設定分類特徵欄位資訊 3. impurity=評估方法(gini or entropy) 4. maxDepth=最大深度   maxBins=每一節點的最大分支數目 | 將線性迴歸的公式轉為sigmoid函式，來界定某個data的類別。透過sigmoid計算出來的p，如果大於0.5，則會歸類為True，否則為False。 |
| Spark ML pipeline | 回歸分析 | DecisionTreeRegressor()  參數：   1. labelCol=標籤欄位 2. featureCol=特徵欄位 |

|  |  |
| --- | --- |
| 演算法 | 介紹 |
| 邏輯回歸 | 簡單迴歸分析  複迴歸 |
| 決策樹 | 利用濕度與氣壓(feature)，以判斷天氣為晴或雨(label) |
| SVM | 支撐向量：一條線描出最大邊緣區，即一條紅線為切割線，兩條平行於切割線並碰到最近的一個點的黑線所圍出來之區域圍最大邊緣區。SVM的目的就是找到這條切割線 |
| 貝氏分析原理 | 利用貝氏定理推導出的以下結果：  推斷在條件之下C發生的相對機率 |
| 隨機森林分類演算法 | 平行產生很多決策樹 |
| GBT梯度提升決策樹 | 一次只產生一棵決策樹，依據前一決策樹的結果，決定如何產生下一決策樹 |