## 项目介绍

用python来编写逻辑回归算法，第三周作业

## 数据介绍

cancer = load\_breast\_cancer()

使用导入的乳腺癌数据，共569个样本，30个特征 阳性：357 阴性：212

## 算法细节

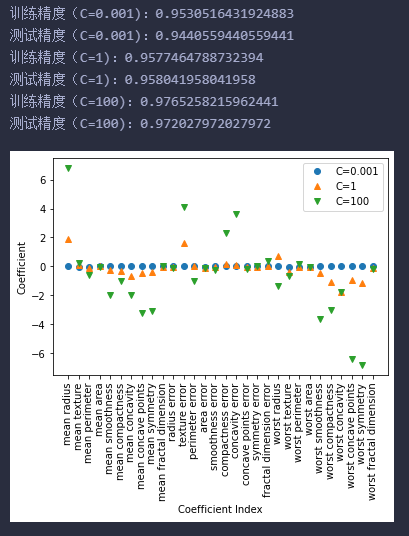
逻辑斯蒂回归（Logistics Regression,LR）又叫逻辑回归或对数几率回归（Logit Regression），是一种用于二分类的线性模型。

损失函数是机器学习里最基础也是最为关键的一个要素，它的作用就是衡量模型预测的好坏。我们对数据进行建模，分别得出了实际模型和预测模型，这两者之间的差距就是损失函数，可以用绝对损失函数来表示。损失函数是很好的反映模型与实际数据差距的工具。

LogisticRegression是一种分类算法，并不是回归算法。在这个公式中我们并没有返回特征的加权求和，而是为预测设置了阈值（0）。如果函数值小于0，我们就预测类别为-1，如果函数值大于0，我们就预测类别+1。对于所有用于分类的线性模型，这个预测规则都是通用的。

　sklearn中LogisticRegression默认使用L2正则化，参数penalty可修改正则化方式。使用sklearn自带的乳腺癌数据集进行逻辑回归训练的，skleran中正则化项C越小，正则化程度越强，参数的变换范围越小。建模时划分列索引为特征值和预测值，并将数据划分成训练集和测试集。训练后，对测试集进行精度评分。最后就是不断调参了。

## 算法结果



IMG_256