1	逻辑回归相比线性回归，有何异同？ 不同之处： 1.**逻辑回归解决的是分类问题，线性回归解决的是回归问题**，这是两者最本质的区别 2.**逻辑回归中因变量是离散的，而线性回归中因变量是连续的**这是两者最大的区别 3 在自变量和超参数确定的情况下**逻辑回归可看作广义的线性模型在因变量下服从二元分布的一个特殊情况** 4.使用最小二乘法求解线性回归时我们认为因变量服从正态分布 相同之处： 1.二者在求解超参数的过程中都使用梯度下降的方法 2.二者都使用了极大似然估计对训练样本进行建模	从作用和数学表达式出发来理解它们之间的异同
2	回归问题常用的性能度量指标 1)**均方误差**：是反映估计值与被估计量之间差异程度的一种度量。 $$\mathrm{MSE}=\frac{\mathrm{SSE}}{\mathrm{n}}=\frac{1}{\mathrm{n}}\sum_{\mathrm{i=1}}^{\mathrm{m}}(f(x_i)-y_i)^2$$ 2)**RMSE 均方根误差**：观测值与真值偏差的平方和与观测次数 m 比值的平方根，用来衡量观测值同真值之间的偏差。 $$\mathrm{RMSE}(\mathrm{X},\mathrm{h})=\sqrt{\frac{1}{\mathrm{m}}\sum_{\mathrm{i=1}}^{\mathrm{m}}(h(x_i)-y_i)^2}$$ 3)**SSE 和方误差** $$\mathrm{SSE}=\frac{1}{\mathrm{n}}\sum_{\mathrm{i=0}}^{\mathrm{n}}(f(x_i)-y_i)^2$$ 4)**MAE**：直接计算模型输出与真实值之间的平均绝对误差 $$\mathrm{MAE}=\frac{1}{\mathrm{n}}\sum_{\mathrm{i=0}}^{\mathrm{n}}\lvert f(x_i)-y_i\rvert$$ 5)**MAPE**：不仅考虑预测值与真实值误差，还考虑了误差与真实值之间的比例。 $$MAPE=\frac{1}{n}\sum_{i=0}^{n}\frac{\lvert f(x_i)-y_i\rvert}{y_i}$$ 6)**平均平方百分比误差** $$MASE=\frac{1}{n}\sum_{i=0}^{n}\left(\frac{\lvert f(x_i)-y_i\rvert}{y_i}\right)^2$$ 7)**决定系数** $$\mathrm{R}^2=1-\frac{\mathrm{SSE}}{\mathrm{SST}}，\text{其中}\mathrm{SST}=\sum_{i=0}^{\mathrm{n}}(y_i-\bar{y})^2$$	归类总结

3

分类问题常用的性能度量指标

常用的性能度量指标有**精确率**、**召回率**、**F_1**、**TPR**、**FPR**。

	预测为真	预测为假
真实为真	TP(true positive)	FN(false negative)
真实为假	FP(false positive)	TN(true negative)

精确率:Precision=TP/(TP+FP). 精确率又称查准率，顾名思义适用于对准确率较高的应用，例如网页检索与推荐。

召回率:Recall=TP/(TP+FN). 召回率又称查全率，适用于检测信贷风险、逃犯信息等。精确率与召回率是一对「**矛盾**」的度量，所以需要找一个「**平衡点**」，往往使用 F_1 是精确率与召回率的调和平均值：

$$\frac{1}{F_1}=\frac{1}{2}\times\frac{1}{P}+\frac{1}{2}\times\frac{1}{R}$$

真正例率:即为正例被判断为正例的概率 TPR=TP/(TP+FN)

假正例率:即为反例被判断为正例的概率 FPR=FP/(TN+FP)

错误率和准确率

错误率：

$$e=\frac{1}{n}\sum_{i=1}^{n}I(f(x_i\neq y_i))$$

准确率：acc=1-e

AUC 与 ROC 曲线

对于 0、1 分类问题，一些分类器得到的结果并不是 0 或 1，如神经网络得到的是 0.5、0.6 等，此时就需要一个「**阈值 cutoff**」，那么小于阈值的归为 0，大于的归为 1，可以得到一个分类结果。**ROC 曲线**(Receiver Operational Characteristic Curve)是以 False Positive Rate 为横坐标，True Postive Rate 为纵坐标绘制的曲线。曲线的点表示了在**敏感度**和**特殊性**之间的平衡，例如越往左，也就是假阳性越小，则真阳性也越小。曲线下面的面积越大，则表示该方法越有利于区分两种类别。**AUC** 即为 ROC 曲线所覆盖的**区域面积**。

归类总结

4	逻辑回归的损失函数 逻辑回归中 sigmoid 函数为 $$h_\theta(x)=\frac{1}{1+e^{(-\theta^T x)}} \qquad \theta^T x=\sum_{i=0}^{n}\theta_i x_i$$ 可以用 **sigmoid 函数表示 0-1 中取 1 的概率**。所以我们的损失函数可以定义为： $$\text{当}y=0\text{时，}Cost(h_\theta(x),y)=-log(1-h_\theta(x))$$ $$\text{当}y=1\text{时，}Cost(h_\theta(x),y)=-log(h_\theta(x))$$ 当我们把损失函数与 0-1 分布的分布律对应起来时： $$p=h_\theta(x)$$ 损失函数就是在 0-1 分布的基础上取对数然后再取负数。损失函数的要求就是预测结果与真实结果越相近，函数值越小，所以会在前面加上负号。当 y=0 时，1-p 的概率会比较大，在前面加上负号，Cost 值就会很小；当 y=1 时，p 的概率会比较大，在前面加上负号，Cost 值就会很小。至于取对数，就是跟最大似然函数有关系，**取对数不影响原本函数的单调性，而且会放大概率之间的差异**，更好的区分各个样本的类别。把上面损失函数写成统一的形式： $$J_{(\theta)}=-\frac{1}{m}\sum_{i=1}^{m}[y^{(i)}log(h_\theta(x^{(i)}))+(1-y^{(i)})log(1-h_\theta(x^{(i)}))]$$	从极大似然函数出发推导
5	逻辑回归处理多标签分类问题时，一般怎么做？ **1. 多项逻辑回归（ Softmax Regression ）** 如果一个样本只对应于一个标签，我们可以假设每个样本属于不同标签的概率服从于几何分布，使用多项逻辑回归（ Softmax Regression ）来进行分类。 $$h_\theta(x)=\begin{bmatrix}p(y=1\mid x;\theta)\\p(y=2\mid x;\theta)\\\vdots\\p(y=k\mid x;\theta)\end{bmatrix}=\frac{1}{\sum_{j=1}^{k}e^{\theta_j^T x}}\begin{bmatrix}e^{\theta_1^T x}\\e^{\theta_2^T x}\\\vdots\\e^{\theta_k^T x}\end{bmatrix}$$ 因此，多项逻辑回归实际上是二分类逻辑回归在多标签分类下的一种拓展 。 **2. k 个逻辑回归分类器** 当存在样本可能居于多个标签的情况时，我们可以训练 k 个二分类的逻辑回归分类器。第 i 个分类器用以区分每个样本是否可以归为第 i 类，训练该分类器时，需要把标签重新整理为“第 i 类标签” 与“非第 i 类标签”两类。不过这样的办法，我们就解决了每个样本可能拥有多个标签的情况。	分类讨论多标签之间是否为有互斥关系