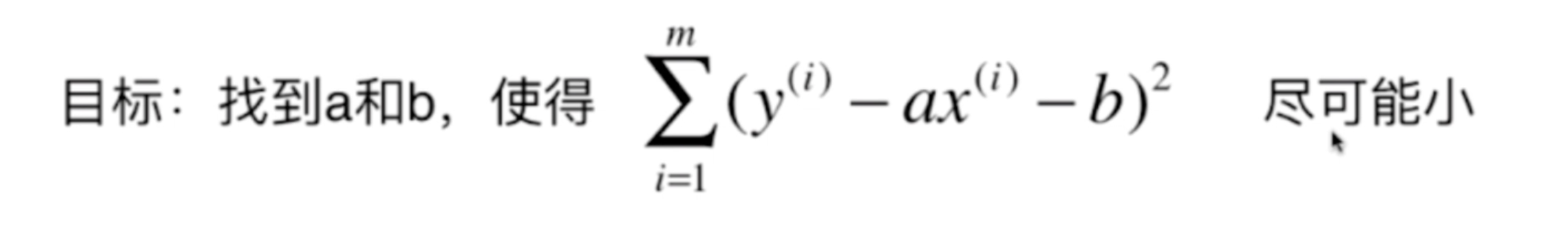
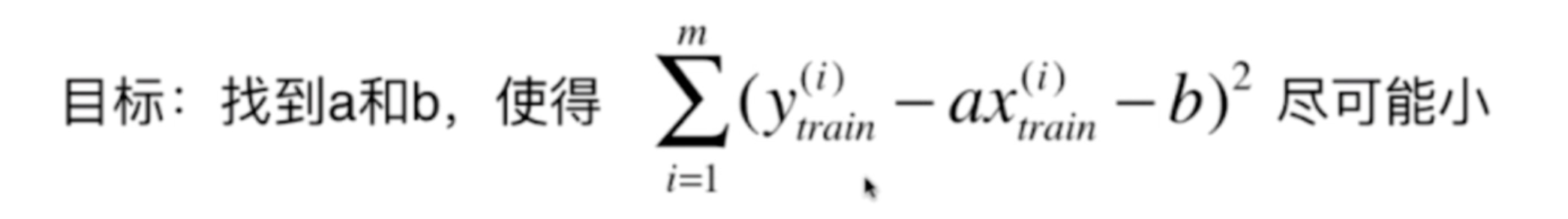
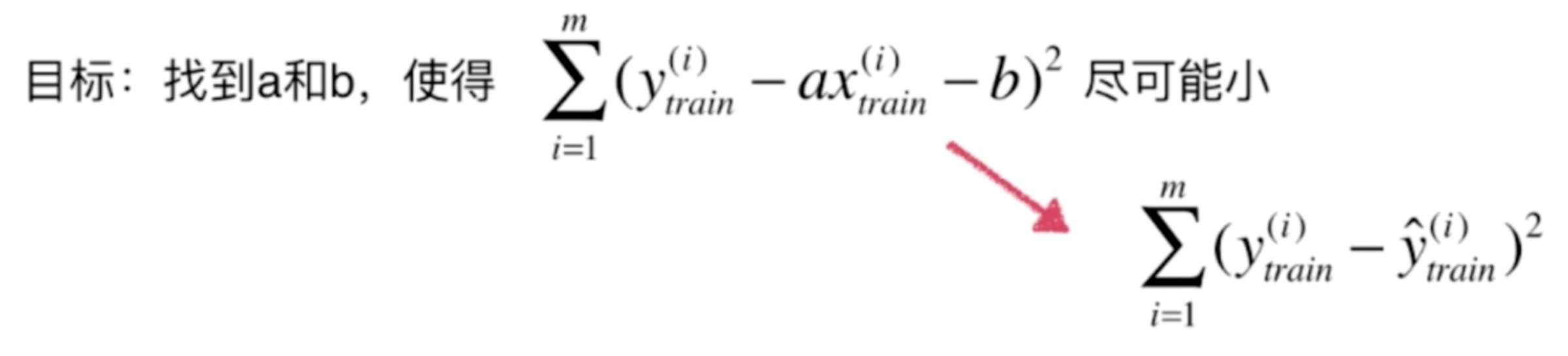
之前的kNN是分类问题，分类问题有分类准确度作为算法模型的评价标准，这里我们来看看回归算法应该用什么作为评价标准，回到之前的损失函数：



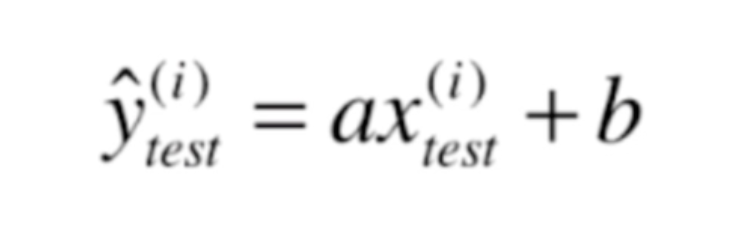
其实在回归算法的使用过程中，也是需要做 train test split的。所以上面的公式可以转换为：



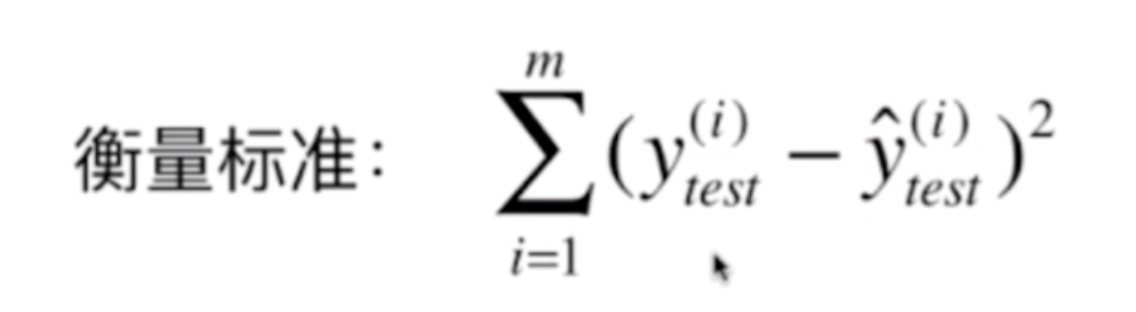
进一步转换：



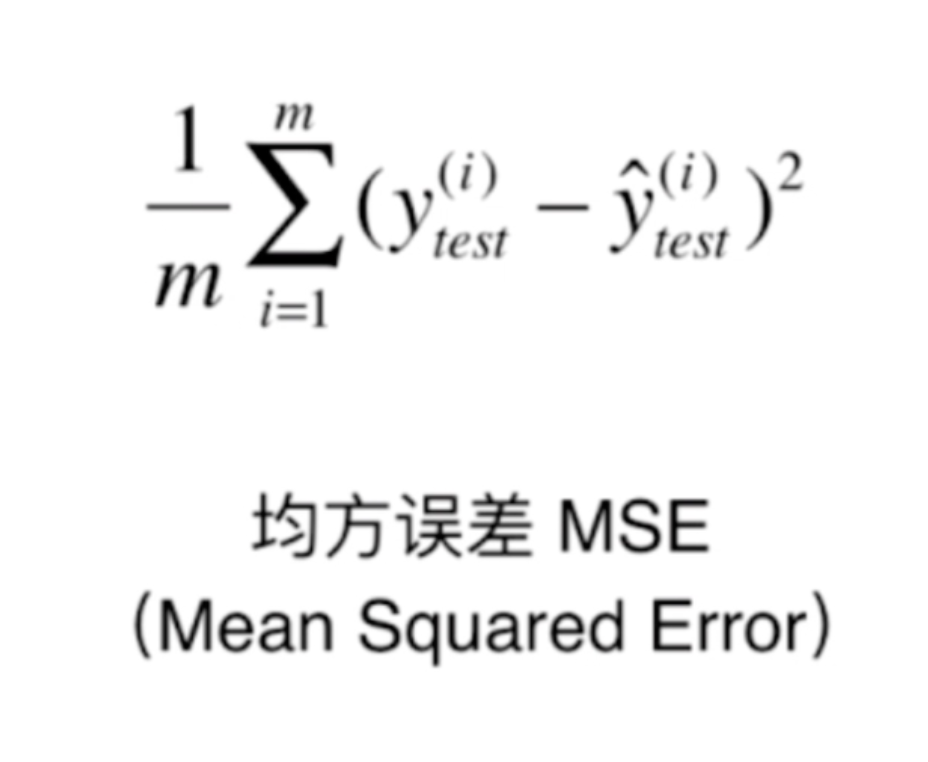
如果fit 出了 a 和 b，就可以算出对应的测试数据的 y\_predict:



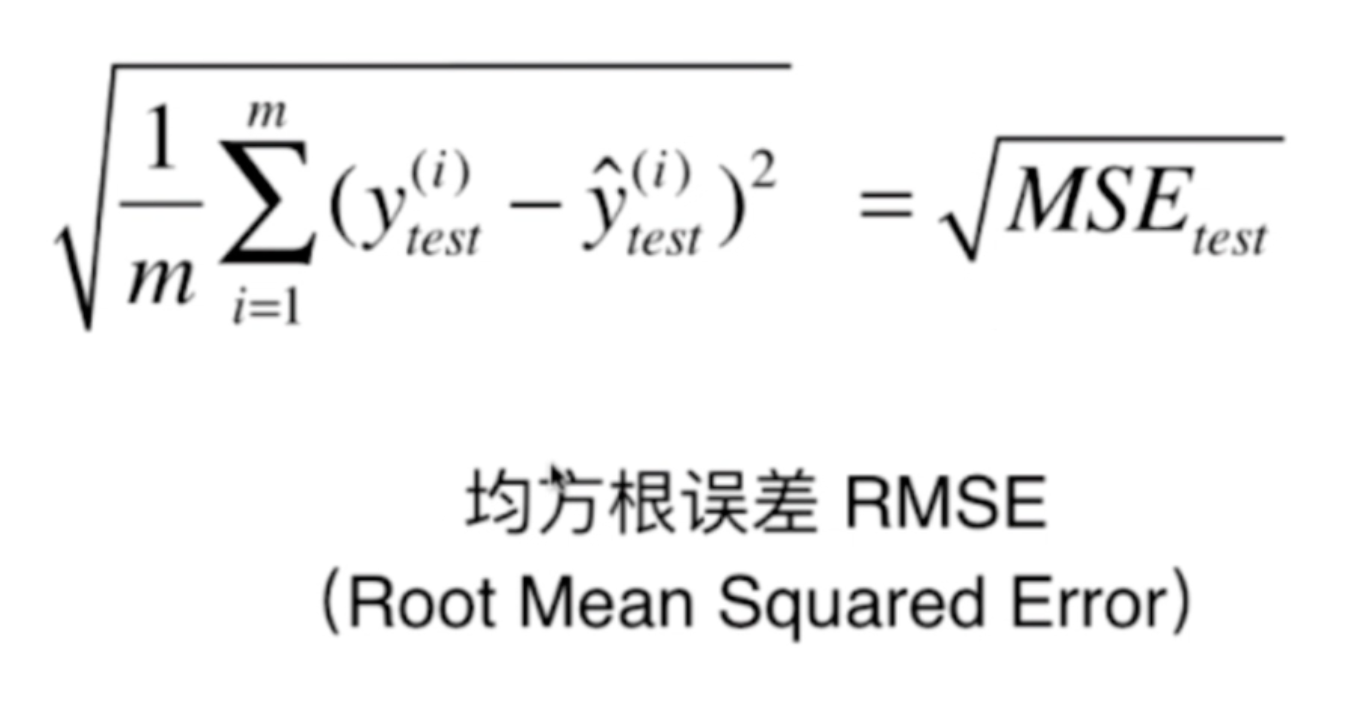
那么标准就是：



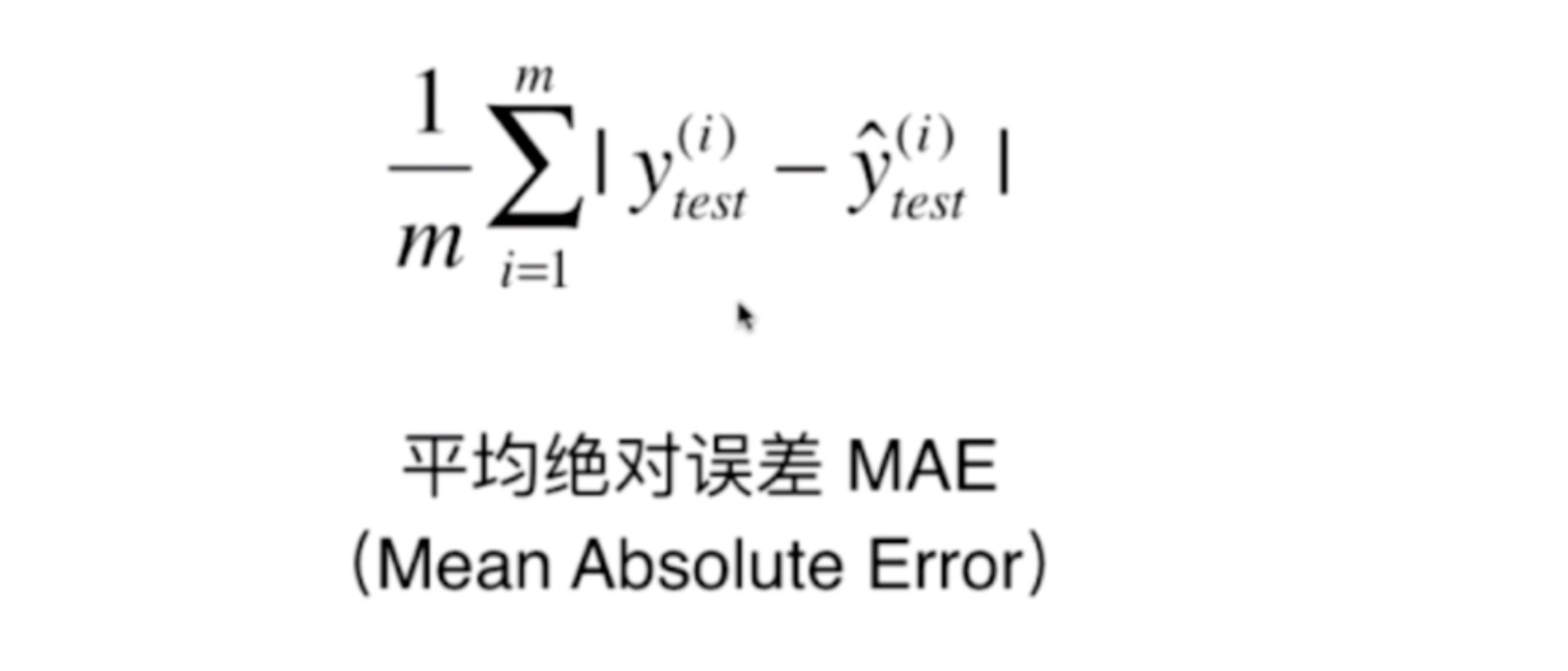
这个还有一个问题，如果数据集的规模（m）变大，上面那个值也会跟着变大，所以要再变换，就得到MSE：



但是上面的公式还是有一个问题： 右边部分是一个二次方的和，左边1/m 是一个一次的常量，所以得到的结果和真实的误差还是有一个二次方的差距，所以又得到了：RMSE



另外，还有一个更加直白的衡量标准：



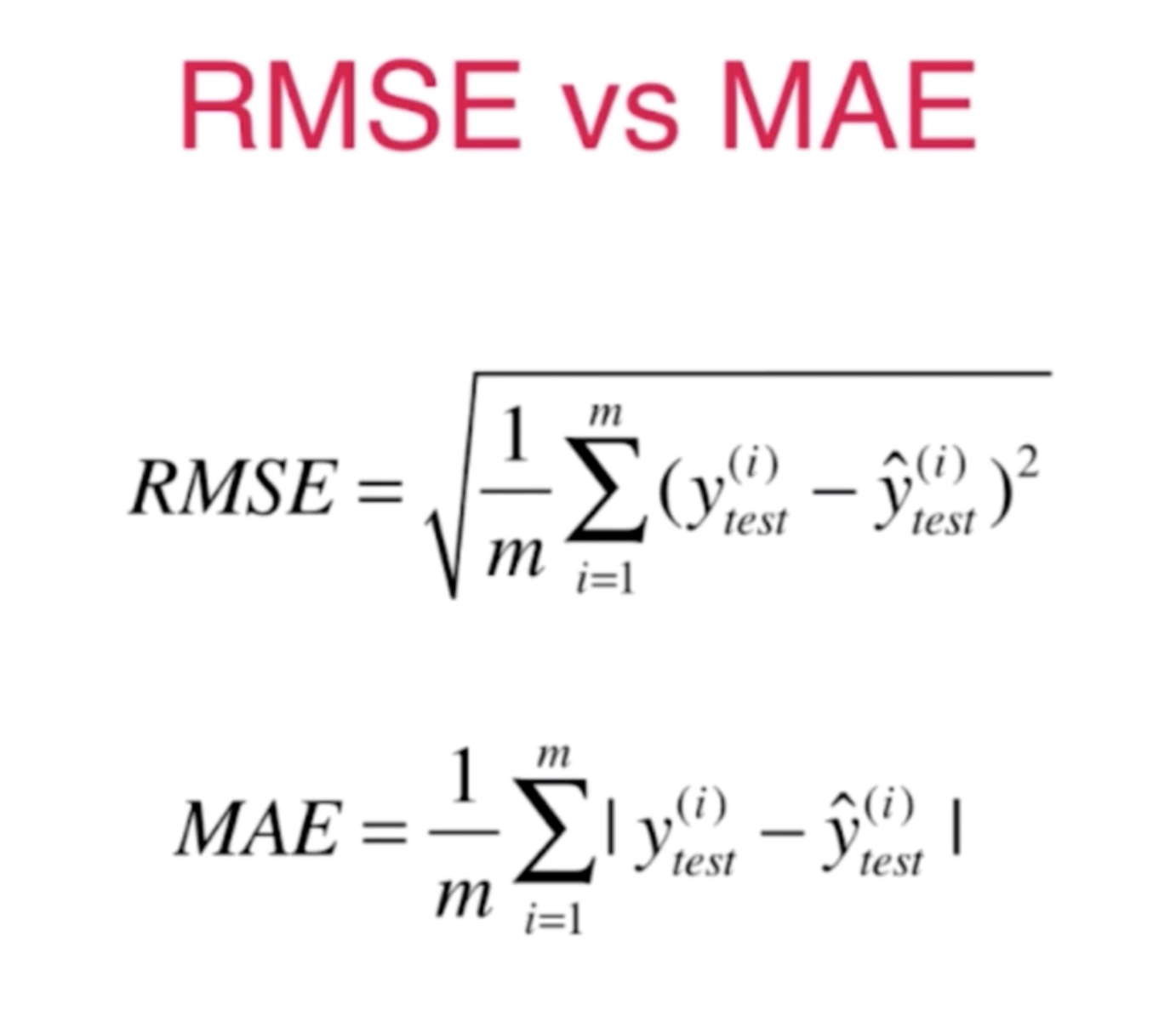
这个带绝对值的表达式，一定很熟悉，因为我们在选线性回归算法的训练损失函数的选择时考虑过这个函数，最终没选它是因为它不是处处可导，不好求导算出极值。但是把它作为评价我们fit出来的模型的方法是很合适的。

在代码中体验上面提到的3中不同的衡量指标：[05-05 Regression-Metrics-MSE-vs-MAE]

最后一个疑问：

为什么 RMSE 和 MAE 都是和 y 一样的量纲，为什么 RMSE 比 MAE 大一些？

看下面的公式



因为 RMSE 中的 平方计算会放大两个值之间的差距

也正是一般 RMSE 大于 MAE，所以我们实际工作中，也用 RMSE 来作为优化的目标，RMSE 越小，训练出来的模型越好。同时RMSE公式中根号下，西格玛中的表达式正好也就是我们选定的损失函数，这样我们训练和评价我们的模型时手段就有了一致性。

End