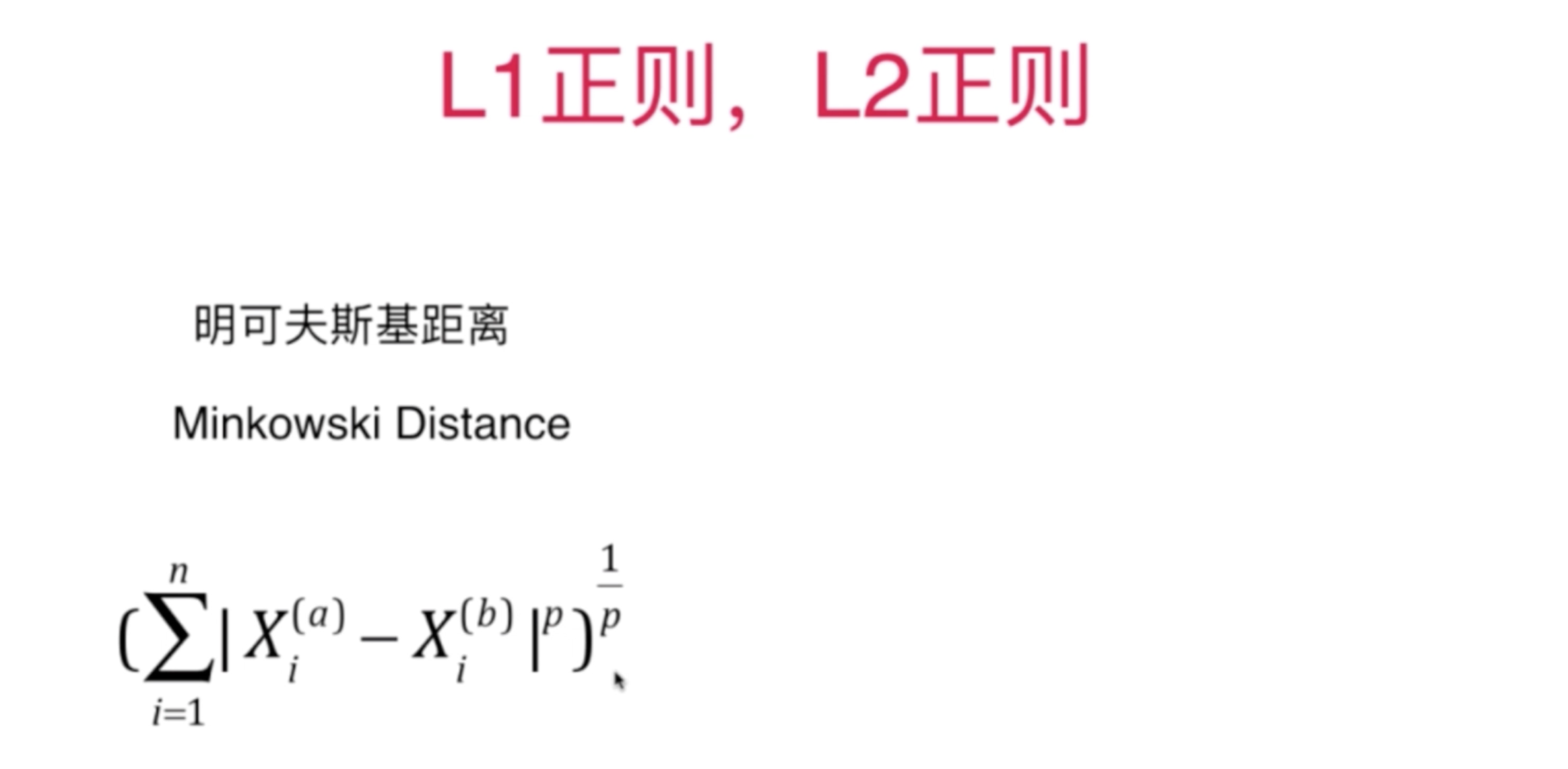
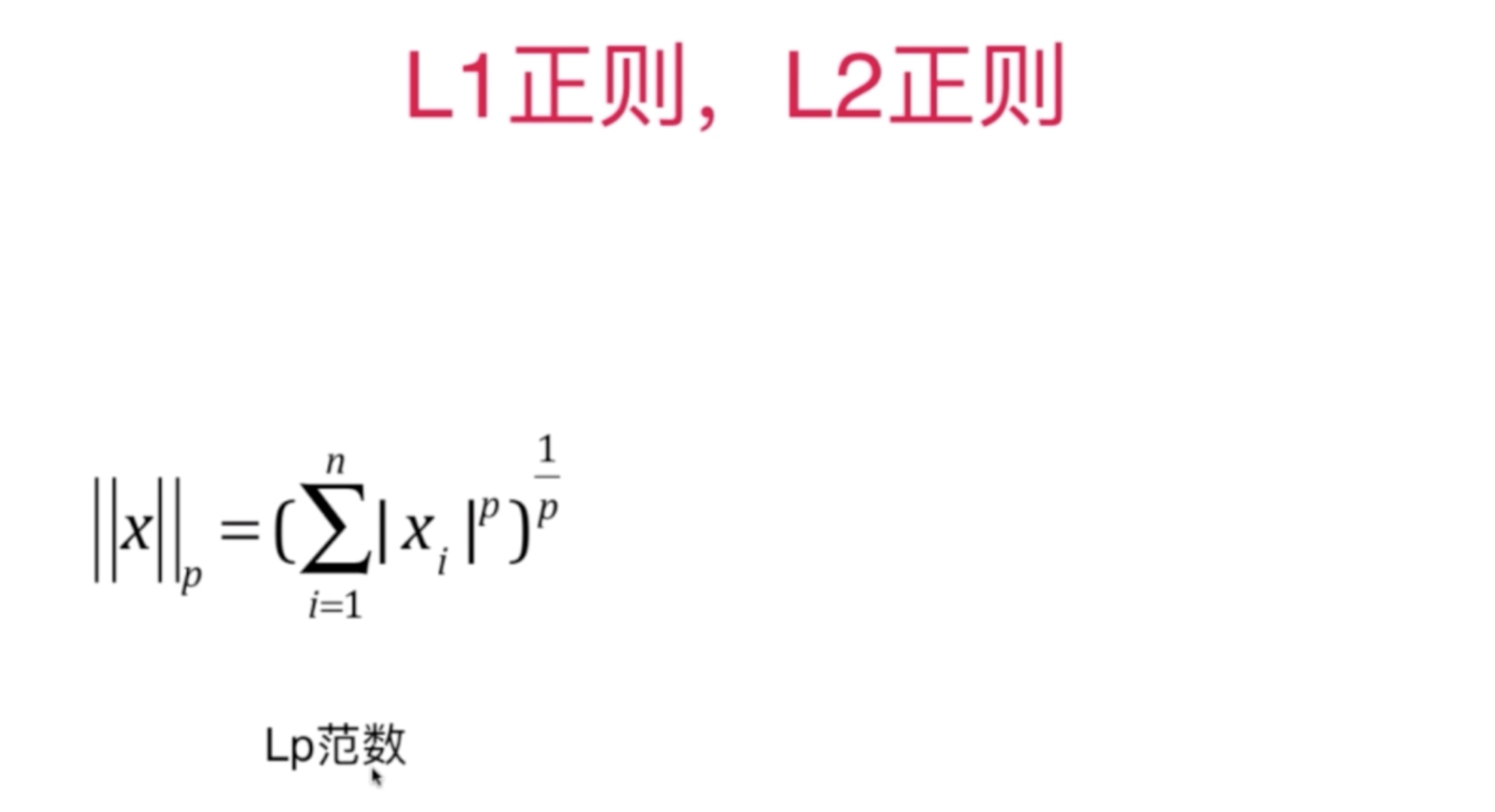
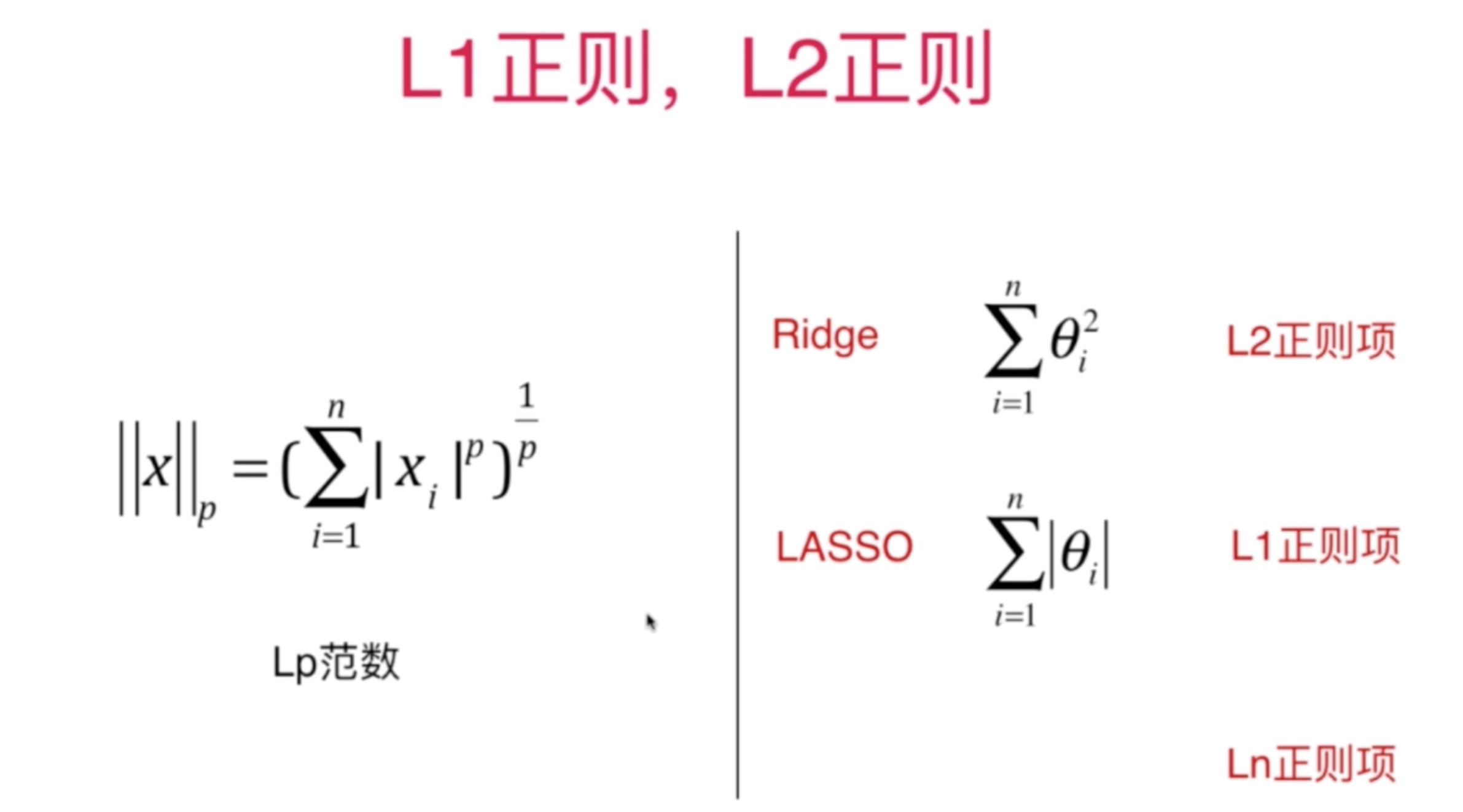


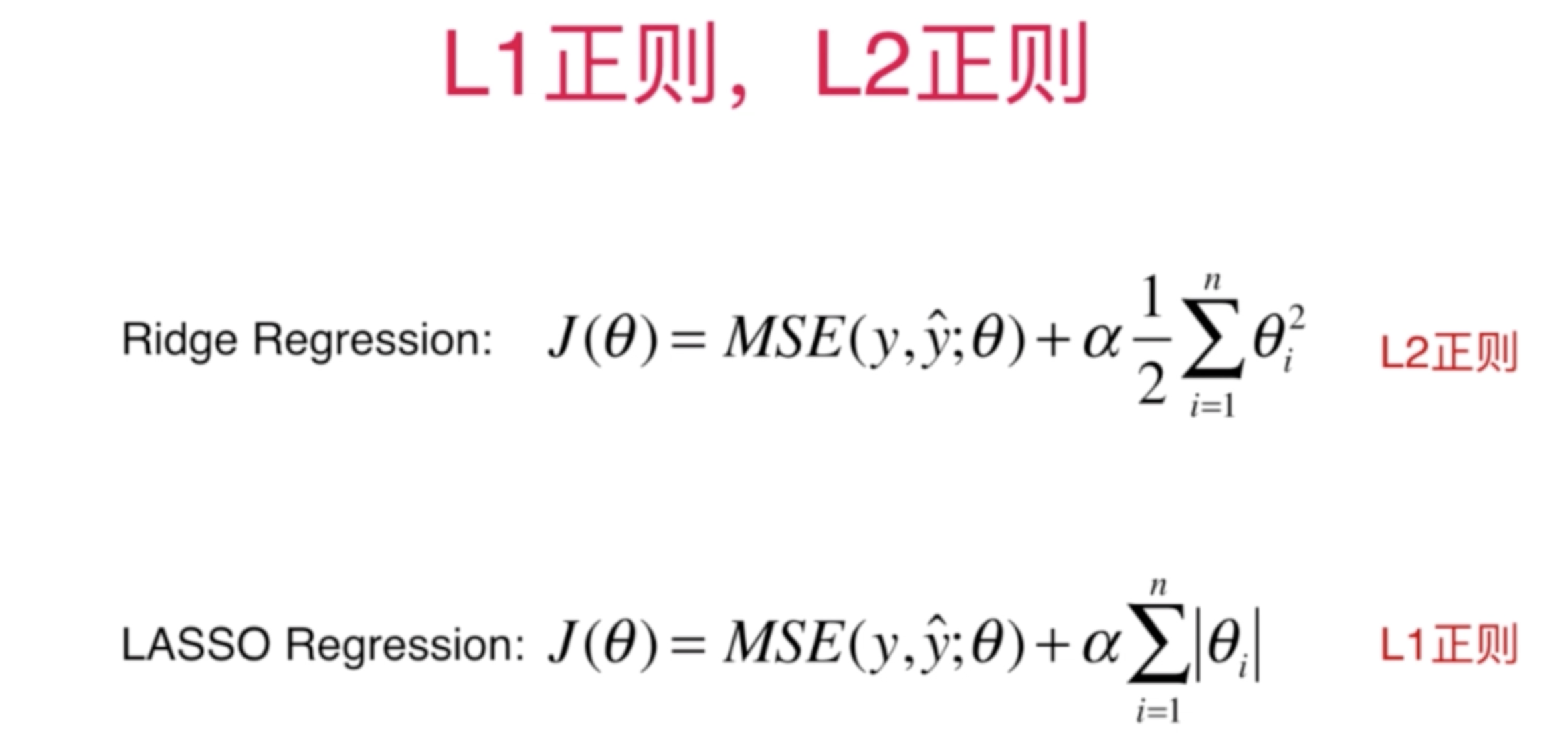
背后用到的数学原理是一样的，

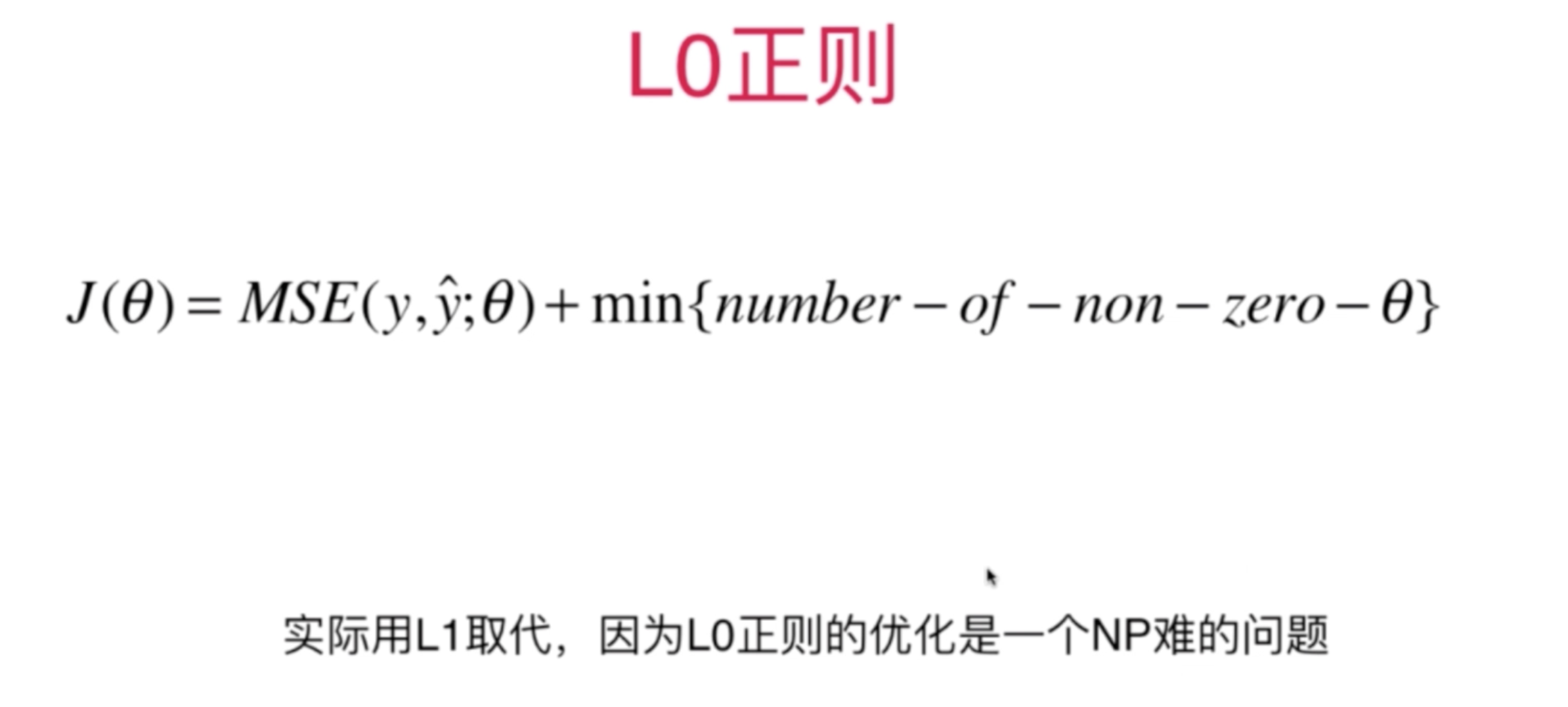


进一步泛化：

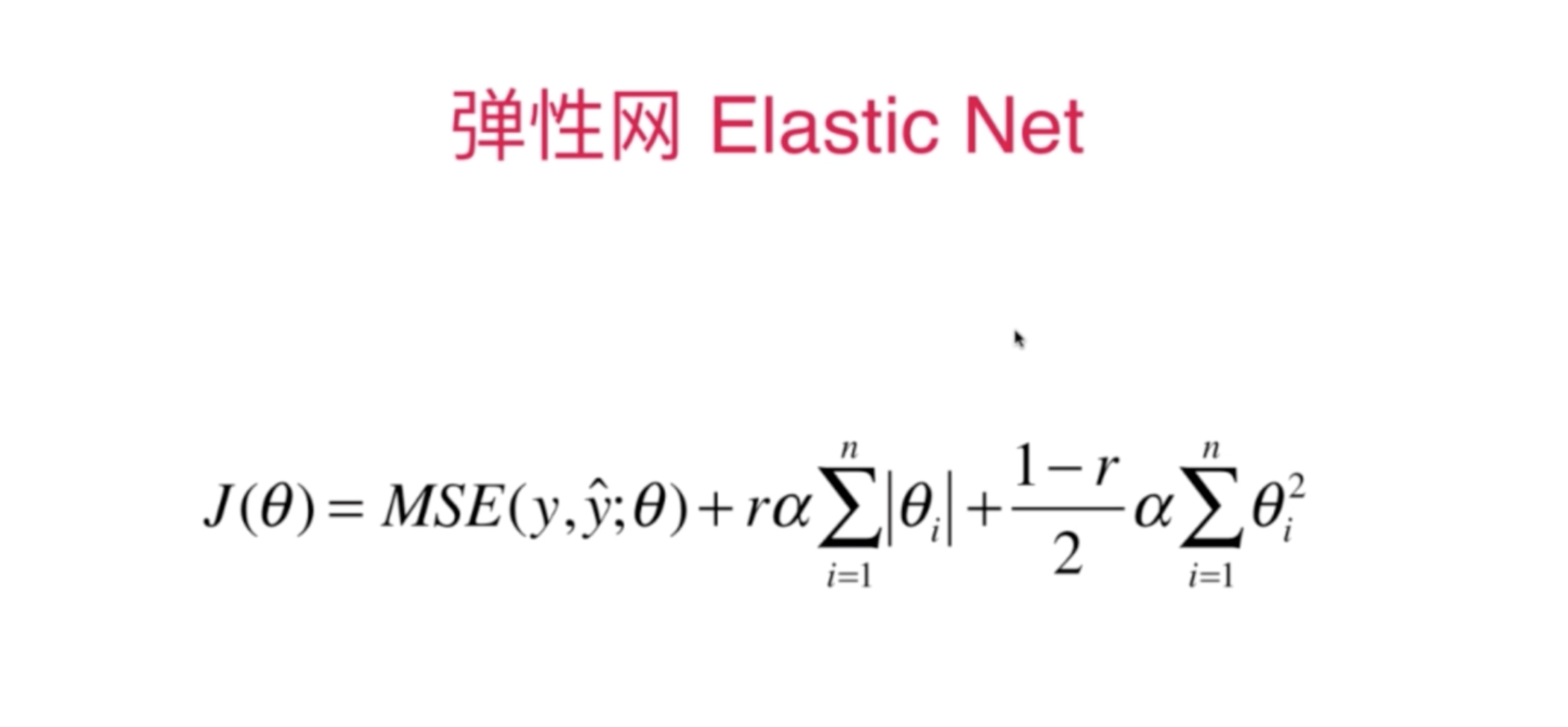








L0正则项的目的是，非零的theta元素个数越少越好，使得曲线不要太陡，那么模型的泛化能力就得以提高。但是L0是一个离散最优化的问题，不能用梯度下降法来求解，所以实际工作中不用。



弹性网是结合了Ridge和LASSO的优势。

实际应用中，对模型正则化的过程中，通常应该先尝试用Ridge回归，Ridge回归的计算相对是精准的，如果计算资源承受得住的话。但是Ridge回归的缺点是如果模型的特征数特别多的话，Ridge不具有特征选择的功能，把能把一些theta设置为0，所以计算量就巨大，此时就应该优先选择弹性网。因为LASSO回归的缺点是它基于将某些theta化为0，其过程中就会产生一些错误，使得我们最终的模型的偏差（Bias）比较大。

就像前面梯度下降中介绍的，小批量下降法是结合了批量下降法和随机批量下降法，将两种方式的优势结合在一起，机器学习领域中经常用这样的方式来创造出新的方法。

模型泛化有点像我们复习考试，平时做的练习题，模拟题，真题就是训练集，验证集，测试集。而最终我们要参加的那场考试就是生产环境下遇到的真实数据集。而模型泛化的最终目的是针对生产环境下的真实数据的预测准确度，就像我们考试复习的最终目的是最后那场考试的得分最高。