机器学习总结

1. 回归

（1）使用极大似然估计解释最小二乘法

误差是独立同分布的，服务均值为0，方差为某定值的高斯分布。

原因：中心极限定理。

（2）中心极限定理的意义：

实际问题中，很多随机的现象可以看作众多因素的独立影响的综合反映，往往近似服从正态分布。

·测试耗电量：大量用户的耗电总和

·测量误差：许多观察不到的、微小误差的总和

·注：应用前提是多个随机变量的和，有些问题是乘性误差，则需要鉴别或者取对数再使用。

（3）似然函数





（4）高斯的对数似然与最小二乘



（5）逻辑回归≡对数线性的

（6）当过拟合，方差变大，将多个方差求平均值，即多个模型集成到一起，可以使得模型变的更好一些。

（7）Lasso具有稀疏性。

（8）MAE绝对值误差平方和，MSE，方差均值，和（6）获得均值一样。即。

（9）epoch说的意思是所有样本都用完一边为一个epoch。32，还是100个赝本是一个batch。

（10）用交叉验证的方式验证几阶好。