**一、线性模型**

**1、共线性数据**

**指模型中的自变量之间存在较强的线性关系**

**容易出现多重共线性的场景：**

**1、数据样本量的不足会导致多重共线性。**

**2、多个变量之间都基于有同趋势的数据。**

**3、多个变量之间存在着近似线性的关系。**

**危害：**

**多重共线性的存在不仅会导致模型的过拟合，而且还会导致回归模型的稳定性和准确性大大的降低。**

**2、岭回归**

**岭回归(英文名：ridge regression, Tikhonov regularization)是一种专用于共线性数据分析的有偏估计回归方法，实质上是一种改良的最小二乘估计法，通过放弃最小二乘法的无偏性，以损失部分信息、降低精度为代价获得回归系数更为符合实际、更可靠的回归方法，对病态数据的拟合要强于最小二乘法。**

**最小二乘法：**

IMG_256

邻回归：

IMG_256

1. **Lasso**

**Lasso 是拟合稀疏系数的线性模型。 它在一些情况下是有用的，因为它倾向于使用具有较少参数值的情况，有效地减少给定解决方案所依赖变量的数量。**

IMG_256

**4.Akaike 信息判据**