

题目：时序数据回归预测

小张是一名制药公司的实验员，最近他们实验室研制了一种新药，能够抑制 A 菌的生长，在多组对照实验中，为了方便监测，小张测量了 A 菌的分泌物 alpha 的含量，其含量能够反映

菌落数的变化。小张经过一番操作：

- (1) 在实验前测量感染了 A 菌的大鼠体内的 alpha 含量；
- (2) 给大鼠注射新药；
- (3) 之后每隔两小时测量一次 alpha；

一共获得了 227 只大鼠的 alpha 变化数据，此时小张已连续工作 24 小时，于是他回去美美地睡了一觉。第二天他来到实验室一看，昨天的大鼠竟然死了近一半，并且记录的数据表上被人不小心洒上了咖啡，经过他的拯救还是有很多数据看不清了。坚强的小张忍痛把剩余的数据整理出来，多年与数据打交道的经验给了他一些直觉，于是他在草稿中写下：

(1) 存活的大鼠的记录一直比较稳定，在用药之后 alpha 一直处于比较低的水平，乍一看 6 小时以后的数据似乎都不高于 0.2（不确定）；

(2) 在死去的大鼠中，alpha 值的变化趋势大致分为两类，一类 alpha 的值在用药后有一个明显的降低，但是很快便再次上升，这是适应能力超强的 A 菌；另一类 alpha 变化的趋势和存活大鼠的趋势十分相似，虽然在用药后 alpha 值下降，但绝对数值都相对要高一些，并且呈微弱的上升趋势，似乎很难与存活小鼠的变化区分；

但这两点都是小张根据直觉写下的内容，并不一定是真实的数据规律，为了知道新药的疗效，

小张只好求助于学习机器学习的同学们：

- (1) 利用机器学习的方法找回缺失的数据；
- (2) 是否能根据补全的数据预测大鼠未来是否能存活呢？
- (3) 怎么才能够知道在小张不存的这段时间内，alpha 是如何变化的呢，构建回归预测模型，根据数据现有的变化规律对后续变化进行预测，预测形式不限。

★ 重要指标

类比前列腺癌中的 PSA 指标

缺失值很多

数据特征

死去的
第二种和
存活类似

→ 插补缺失值