

9. 在 categorical 预测中.

intercept: 是 y 分类为第 1 类别的值.

slope: 平均变化在 y 上 between the two categories.

10. 多元线性回归

a. $R^2_{adj} = 1 - \left(\frac{SSE}{SST} \times \frac{n-1}{n-k-1} \right)$

i. $k \geq 0 \Rightarrow R^2_{adj} \leq R^2$

ii. R^2_{adj} 也在用了对加入模型中变量个数进行惩罚

iii. 我们选择高 R^2_{adj} 的模型.

b. model selection

i. ~~stepwise model~~ backward elimination: 从全模型(所有特征都保留), 每次剔除一个特征

ii. forward elimination: 从空 model (没有特征), 每次添加一个特征.

iii. 标准: p-value, adjust R^2 .

3.3

IV. backward elimination — adjusted R^2

start with full model

去掉一个特征, 计算 adjust R^2 .

选择当前为止最高 adjust R^2 的 model.

Repeat until adjust R^2 不再增大

V. backward elimination — p-value

full model

去掉 p-value 最大的特征.

repeat until 所有剩下的变量 significant

i. adjust R^2 v.s. p-value.

p-value: significant 力

adjusted R^2 : 更可靠的力

p-value 方法依赖于 5% significant level cutoff









