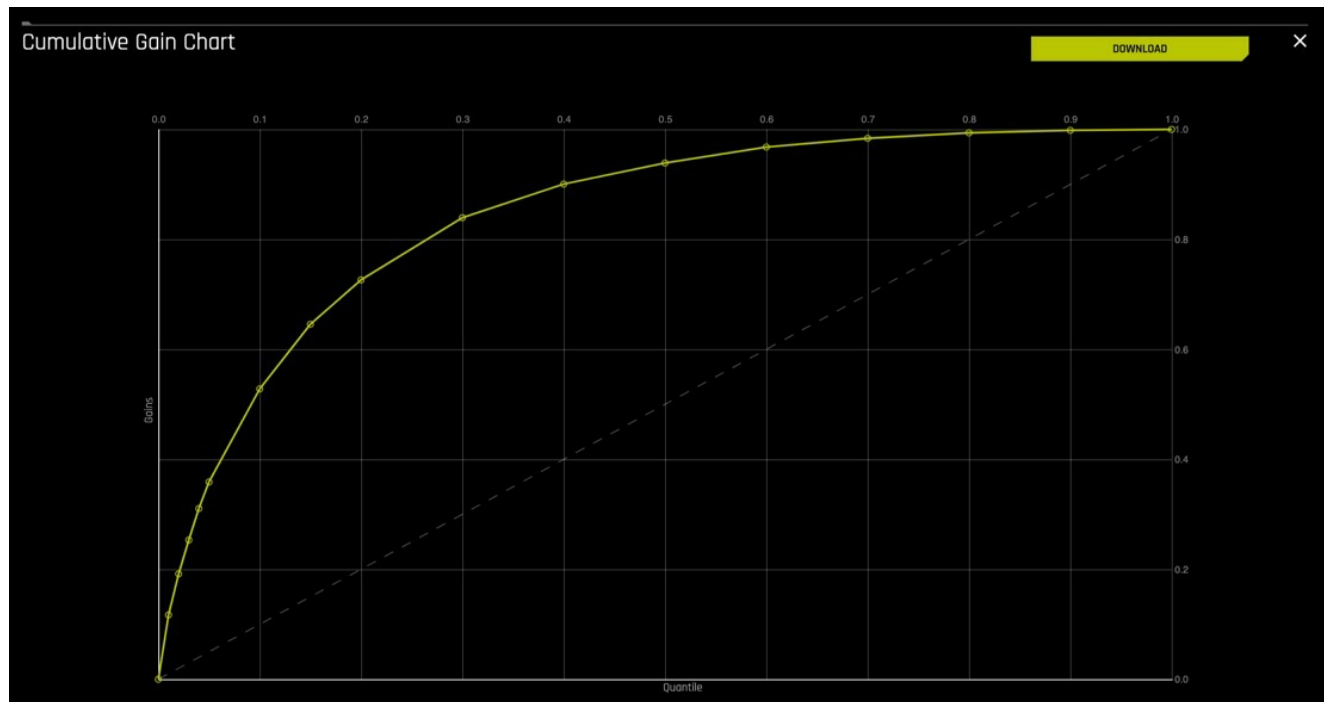


机器学习实验评分和分析教程-财务重点

h2oai.github.io/tutorials/machine-learning-experiment-scoring-and-analysis-tutorial-financial-focus

10.任务8：ER：收益

继续在诊断页面上，选择“**累积增益**”曲线。增益曲线应类似于以下曲线：



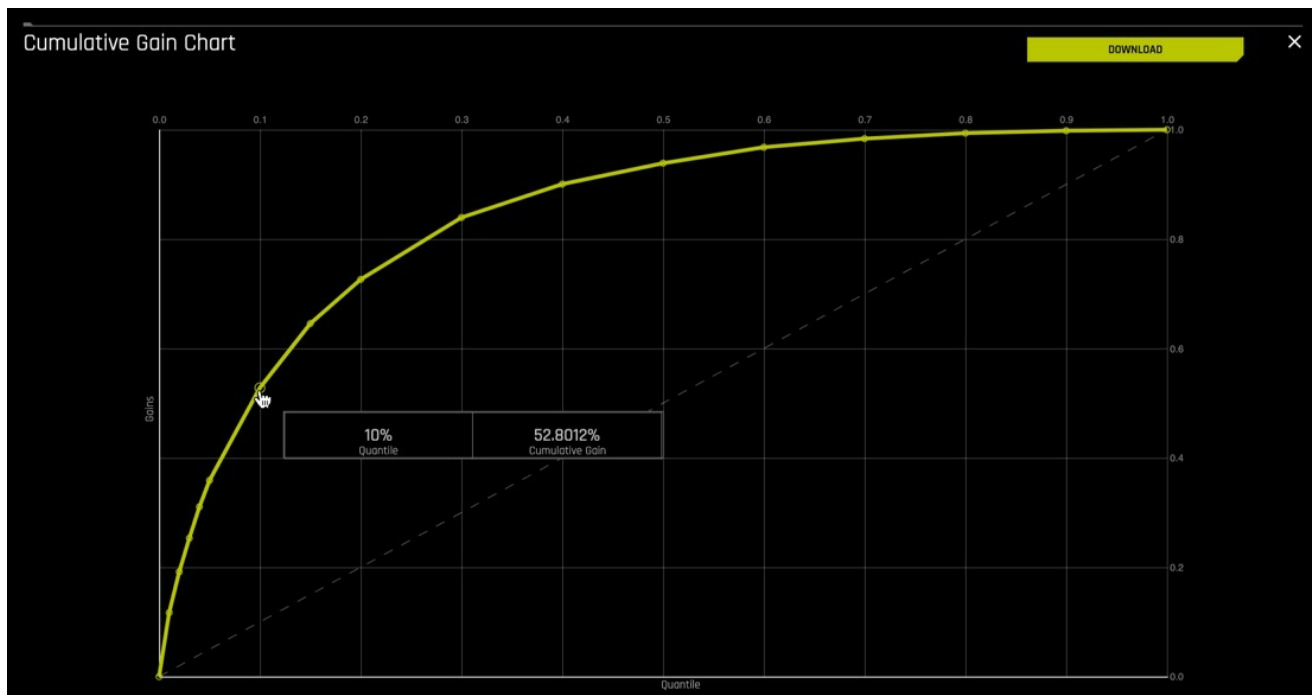
请记住，对于**增益**曲线：

- 累积增益图是衡量模型性能的直观工具。
- y轴显示阳性反应的百分比。这是所有可能的积极回应中的百分比
- x轴显示的是Freddie Mac数据集中所有未违约的客户所占的百分比，仅占全部案例的一小部分
- 虚线是基线（总体响应率）
- 它有助于回答以下问题：“在积极目标类别的所有观察中，有百分之几是预测的最高百分比为1%，2%，10%等（累积）？”根据定义，100%的增益为1.0。

注意：已对图的y轴进行了调整，以表示分位数，这使您可以专注于数据量最大，影响最大的分位数。

1.将鼠标悬停在增益图表上的各个分位数上，以查看分位数百分比和累积增益值

2. 1%，2%，10%的分位数的累积增益是多少？



For this Gain Chart, if we look at the top 1% of the data, the at-chance model (the dotted diagonal line) tells us that we would have correctly identified 1% of the defaulted mortgage cases. The model generated (yellow curve) shows that it was able to identify about 12% of the defaulted mortgage cases.

If we hover over to the top 10% of the data, the at-chance model (the dotted diagonal line) tells us that we would have correctly identified 10% of the defaulted mortgage cases. The model generated (yellow curve) says that it was able to identify about 53% of the defaulted mortgage cases.

3. Based on the shape of the gain curve and the baseline (white diagonal dashed line) would you consider this a good model?

请记住，理想的预测模型起初非常陡峭，并且根据经验法则，曲线越陡峭，增益越高。基线（白色对角虚线）和增益曲线（黄色曲线）之间的区域（众所周知，曲线下的区域）在视觉上向我们展示了我们的模型比随机模型好多少。总有改进的余地。增益曲线可能更陡峭。

注意：如果不确定什么是AUC或什么是增益图，请随时阅读本教程的概念部分。

4.通过单击图右上角“**下载**”选项旁边的x退出增益图表。

更深入的潜水和资源

[迈向数据科学-理解和解释增益图和提升图](#)

[背部下一个](#)

