

# HOMEWORK 1

1. 已知针对模型  $f$ ，使用训练数据集得到的估计记为  $\hat{f}$ ，现有独立于训练数据集的  $(x_0, y_0)$ ， $y_0 = f(x_0) + \epsilon$ ，其中  $\epsilon$  为随机误差项，证明：

$$\mathbb{E} \left( y_0 - \hat{f}(x_0) \right)^2 = \text{Var} \left( \hat{f}(x_0) \right) + \left[ \text{Bias} \left( \hat{f}(x_0) \right) \right]^2 + \text{Var}(\epsilon).$$

2. 请简要叙述一个可以应用数据挖掘的分析场景。请具体说明：（1）该分析场景中需要解决的业务问题；（2）针对该问题需要收集的数据，以及该数据中的因变量和自变量信息。

本次作业以 PDF 的形式提交，需要包括详细的证明步骤。

**截止时间：3 月 16 日 20:00。请预留一定的时间，迟交作业扣 3 分，作业抄袭 0 分。**