

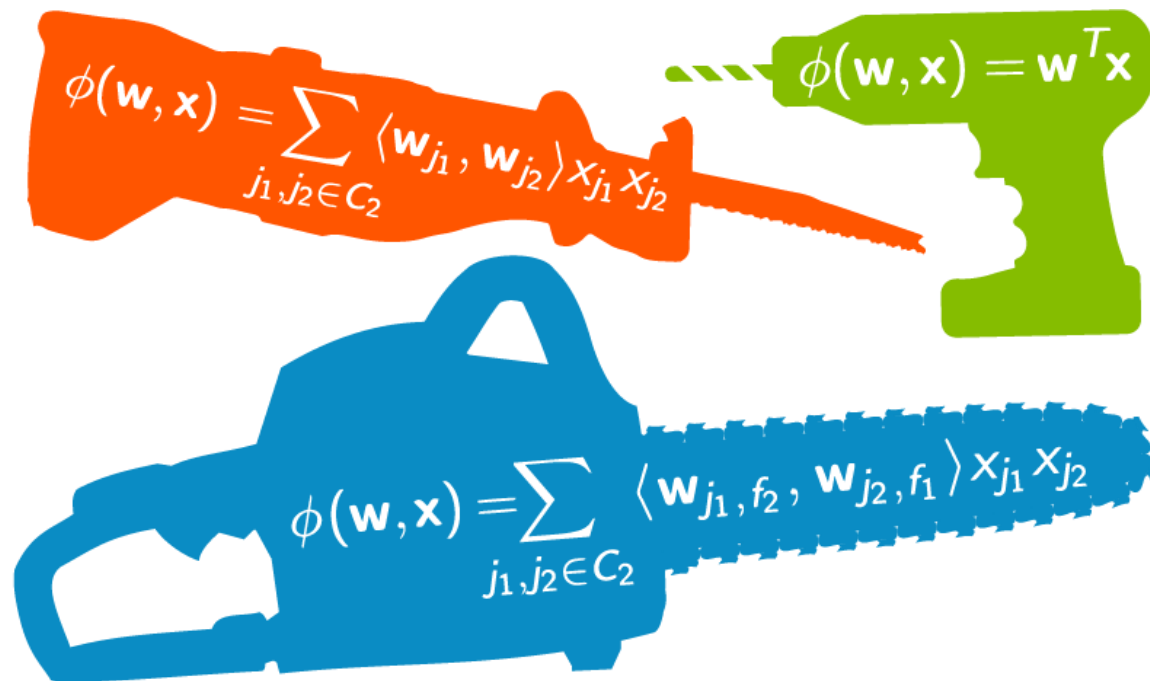
# 法律声明

---

- 本课件包括：演示文稿，示例，代码，题库，视频和声音等，小象学院拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意，我们将保留一切通过法律手段追究违反者的权利。



关注 **小象学院**



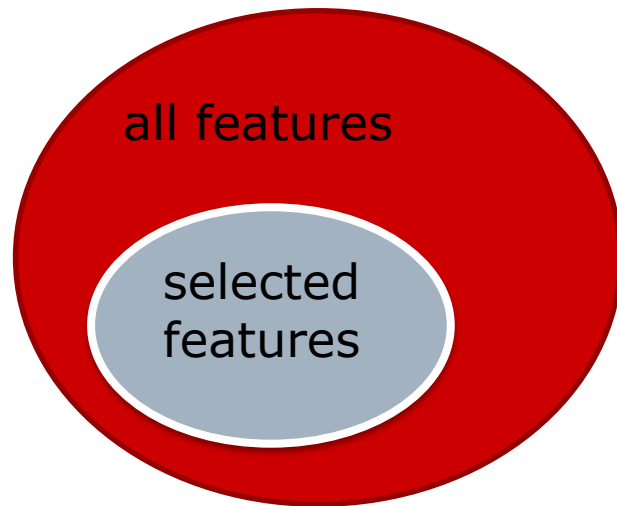
## 特征工程

--Robin

# 特征选择

---

- 功能：
  - 减少特征数量、降维，使模型泛化能力更强，减少过拟合
  - 增强特征和特征值之间的理解
- 最终得到的特征选是原来特征的一个子集。
- 方法：
  1. 去掉取值变化小的特征
  2. 基于单变量统计特征选择
  3. 基于模型的特征选择



# 特征选择

---

- 去掉取值变化小的特征
  - 极端情况，如果所有样本在某个维度上的特征全都相同，即0方差，说明该特征描述或代表样本的能力很弱
  - `sklearn.feature_selection.VarianceThreshold`
- 基于单变量统计特征选择
  - 对每一个特征进行测试，衡量该特征和标签之间的关系
  - `sklearn.feature_selection.SelectKBest`，移除得分前 k 名以外的所有特征
- 基于模型的特征选择
  - 根据特征的 `importance_score` 排名选择，如：Decision Tree
  - `sklearn.feature_selection.SelectFromModel`

# 联系我们

---

小象学院：互联网新技术在线教育领航者

— 微信公众号：**小象学院**

