

Python 3 玩儿转机器学习

讲师：liuyubobobo

版权所有 侵权必究
liuyubobobo

慕课网《Python3机器学习》

集成学习

讲师：liuyuboboo

版权所有，侵权必究

慕课网《Python3机器学习》

什么是集成学习

讲师：liuyuboboy

版权所有，侵权必究

什么是集成学习



什么是集成学习



新世纪福音战士 (EVA)

超级电脑 MAGI

三贤者

母亲的身份；科学家的身份；女性的身份

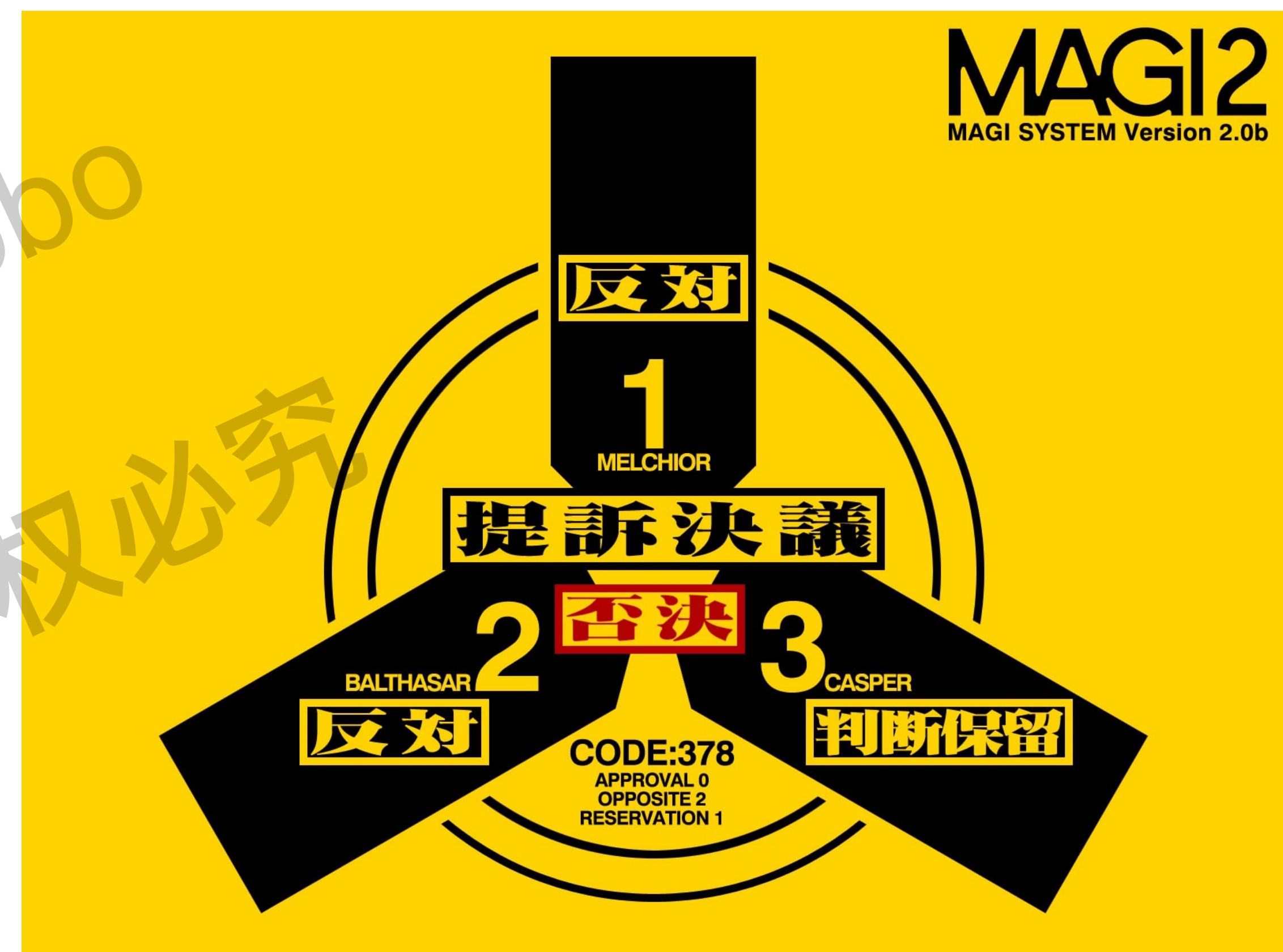
什么是集成学习

新世纪福音战士 (EVA)

超级电脑 MAGI

三贤者

母亲的身份；科学家的身份；女性的身份



什么是集成学习

生活中的集成学习：

买东西找别人推荐？

病情确诊？

什么是集成学习

kNN

逻辑回归

SVM

决策树

神经网络

贝叶斯



投票；少数服从多数

Voting Classifier

实践：scikit-learn中的Voting Classifier

慕课网《Python3机器学习》
讲师：liuyubobobo
版权所有，侵权必究

慕课网《Python3机器学习》

Soft Voting

讲师：liuyubobobo

版权所有，侵权必究

Voting Classifier

更合理的投票，应该有权值

唱歌比赛的投票？

Voting Classifier

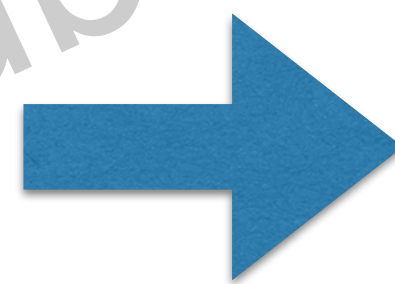
模型1 A-99%； B-1%

模型2 A-49%； B-51%

模型3 A-40%； B-60%

模型4 A-90%； B-10%

模型5 A-30%； B-70%



A-两票； B-三票

最终结果为B

Hard Voting

Soft Voting

模型1 A-99%； B-1%

$$A - (0.99 + 0.49 + 0.4 + 0.9 + 0.3) / 5$$

模型2 A-49%； B-51%

$$= 0.616$$

模型3 A-40%； B-60%

$$B - (0.01 + 0.51 + 0.6 + 0.1 + 0.7) / 5$$

模型4 A-90%； B-10%

$$= 0.384$$

模型5 A-30%； B-70%

最终结果为A

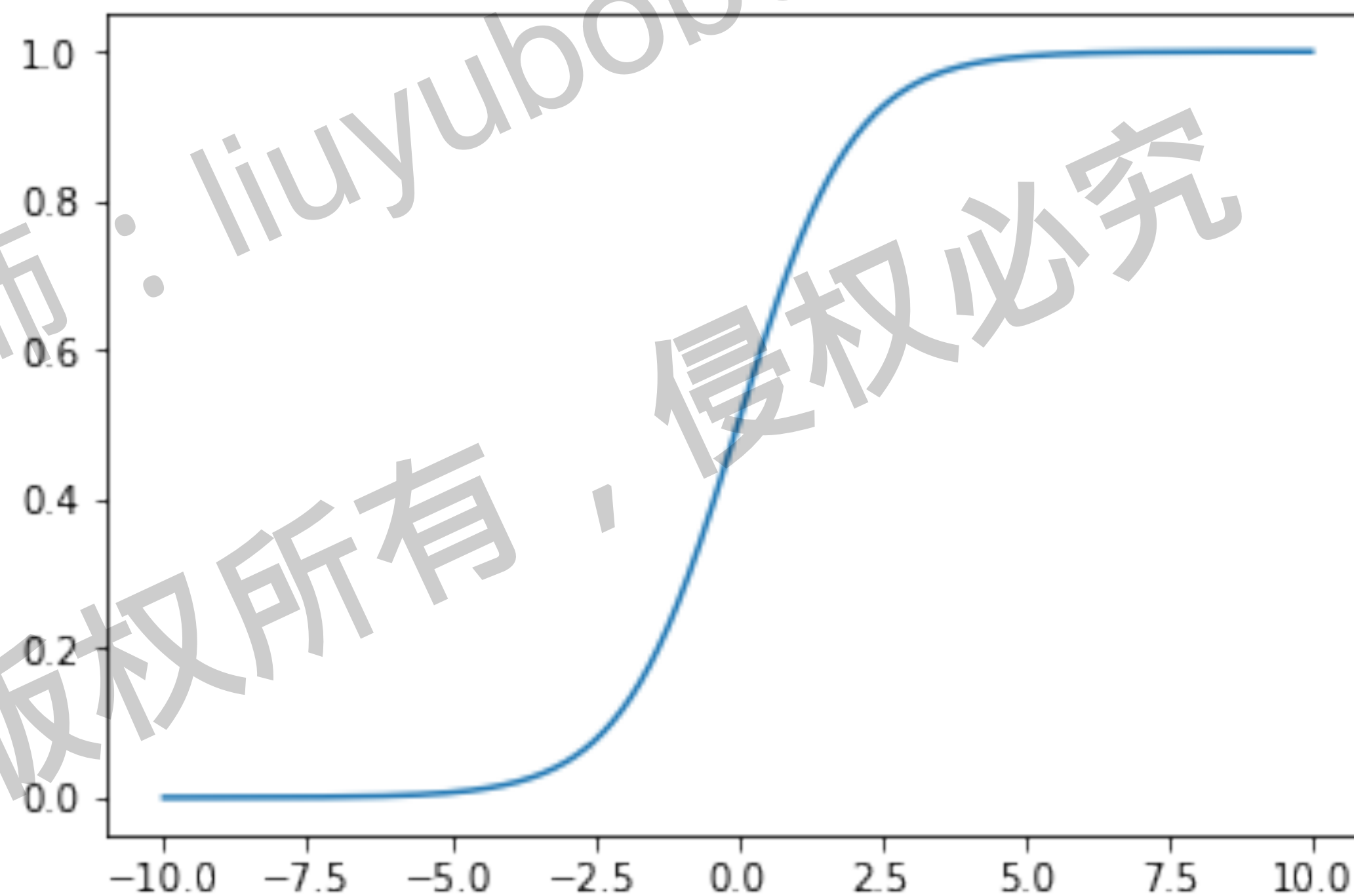
Soft Voting

要求集合的每一个模型都能估计概率

`predict_proba`

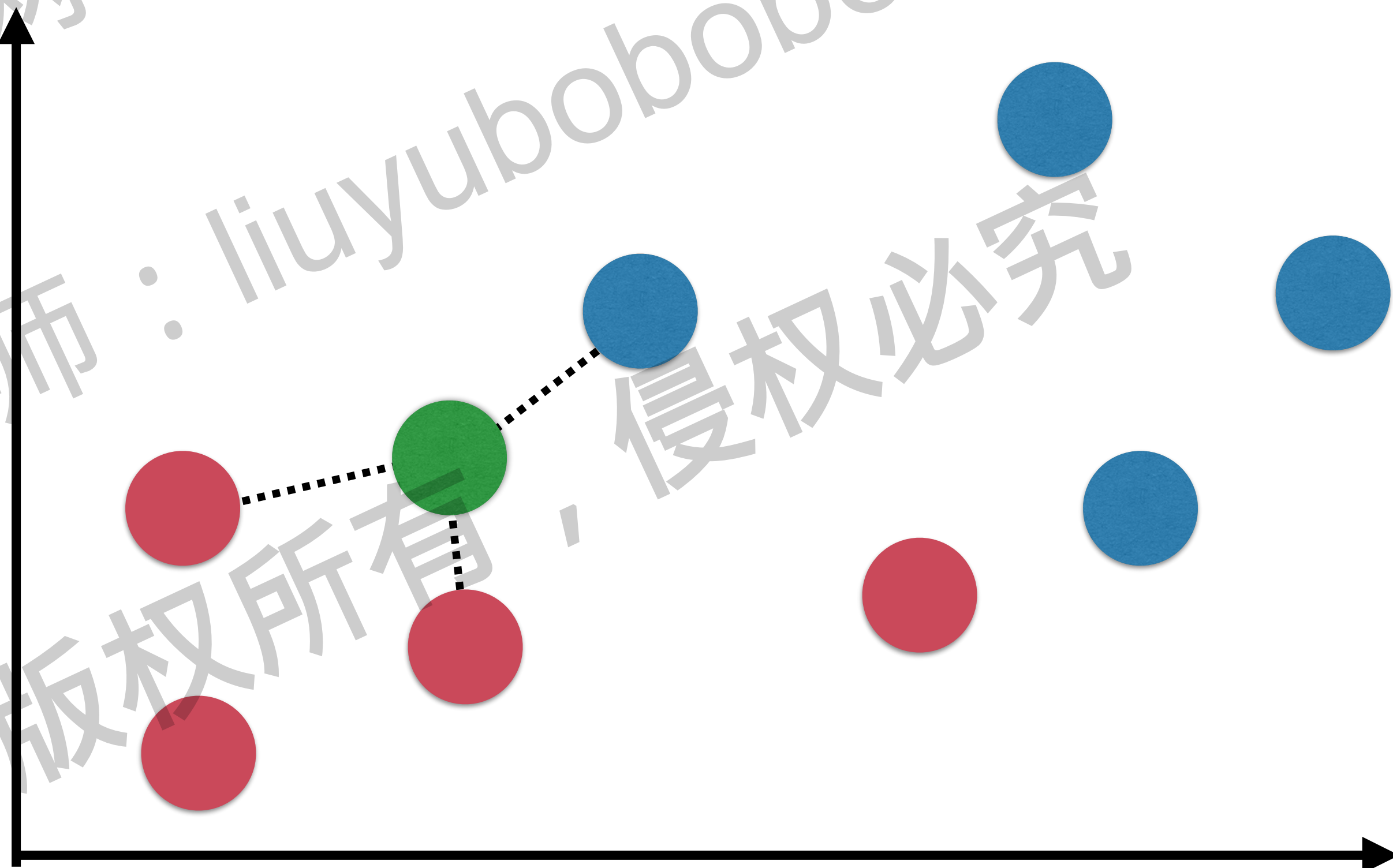
Soft Voting

逻辑回归 本身就是基于概率模型的



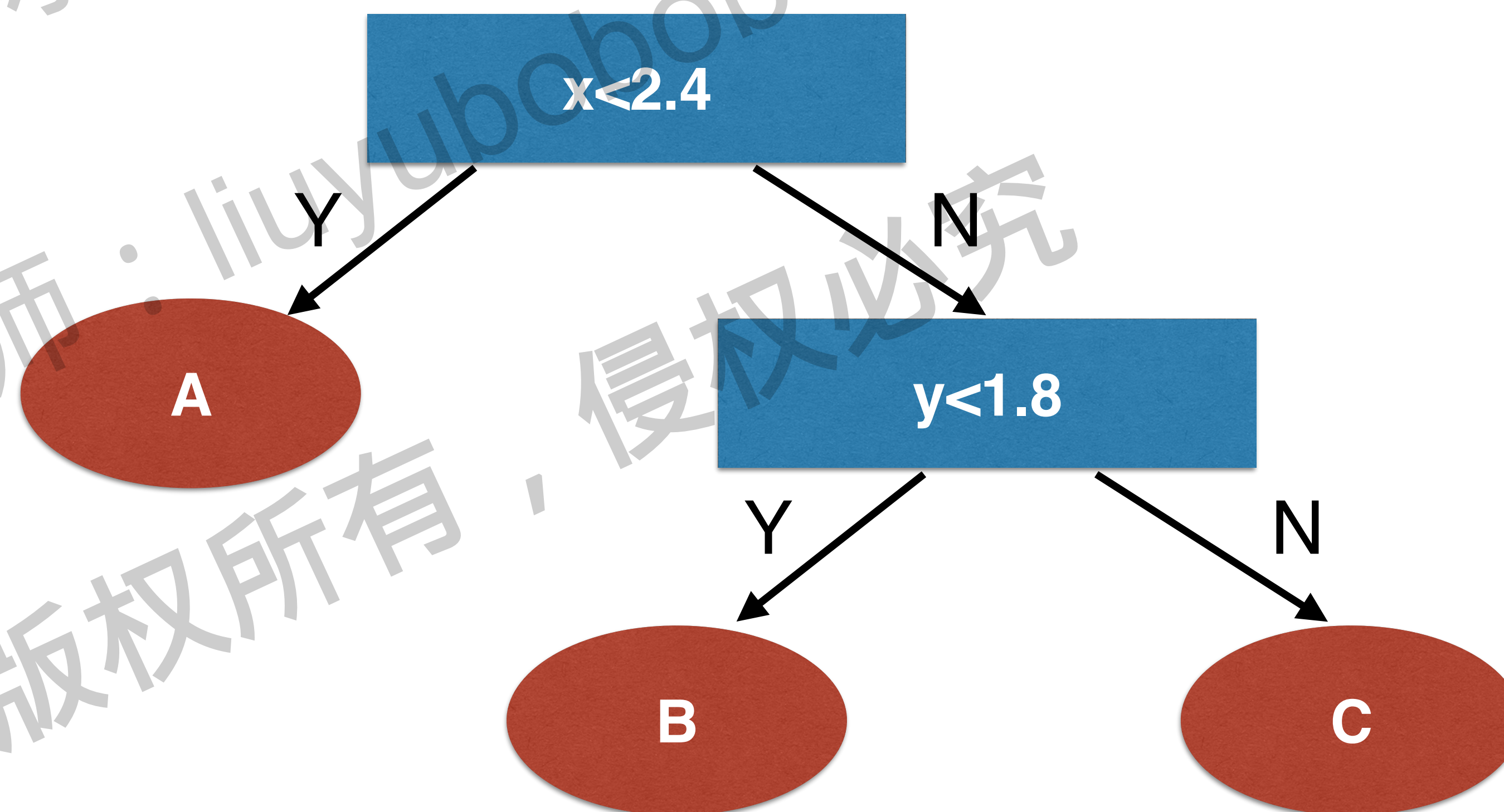
Soft Voting

kNN



Soft Voting

决策树



Soft Voting

SVC

probability : boolean, optional (default=False)

Whether to enable probability estimates. This must be enabled prior to calling fit, and will slow down that method.

<http://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>

慕课网《Python3机器学习》

实践：Soft Voting

讲师：liuyuboboo

版权所有，侵权必究

慕课网《Python3机器学习》

Bagging和Pasting

讲师：liuyuboboo

版权所有，侵权必究

集成学习

虽然有很多机器学习方法，

但是从投票的角度看，仍然不够多

创建更多的子模型！集成更多的子模型的意见。

子模型之间不能一致！子模型之间要有差异性。

集成学习

创建更多的子模型！ 集成更多的子模型的意见。

子模型之间不能一致！ 子模型之间要有差异性。

如何创建差异性？

每个子模型只看样本数据的一部分。

集成学习

如何创建差异性？

每个子模型只看样本数据的一部分。

例如：一共有500个样本数据；每个子模型只看100个样本数据

每个子模型不需要太高的准确率

集成学习

每个子模型不需要太高的准确率

如果每个子模型只有51%的准确率

如果我们只有1个子模型，整体准确率：51%

如果我们只有3个子模型，整体准确率： $0.51^3 + C_3^2 \cdot 0.51^2 \cdot 0.49$ 51.5%

如果我们只有500个子模型，整体准确率： $\sum_{i=251}^{500} C_{500}^i \cdot 0.51^i \cdot 0.49^{500-i}$ 65.6%

集成学习

每个子模型不需要太高的准确率

如果每个子模型只有60%的准确率

如果每个模型的准确率是60%：
$$\sum_{i=251}^{500} C_{500}^i \cdot 0.6^i \cdot 0.4^{500-i}$$
 99.999%

Bagging 和 Pasting

如何创建差异性？

每个子模型只看样本数据的一部分。

取样：放回取样，不放回取样

放回取样： Bagging

不放回取样： Pasting

Bagging 更常用

Bagging 和 Pasting

统计学中，放回取样：bootstrap

实践：Bagging 和 Pasting

讲师：liuyubobobo
版权所有，侵权必究

更多和Bagging相关的讨论

讲师：liuyuboboo
版权所有，侵权必究

OOB Out-of-Bag

放回取样导致一部分样本很有可能没有取到

平均大约有37%的样本没有取到。

不使用测试数据集，而使用这部分没有取到的样本做测试 / 验证。

`oob_score_`

慕课网《Python3机器学习》

实践：oob

讲师：liuyubobobo

版权所有，侵权必究

Bagging的更多探讨

Bagging的思路极易并行化处理

n_jobs

实践：bagging并行化处理

讲师：liuyuboboo
版权所有，侵权必究

Bagging的更多探讨

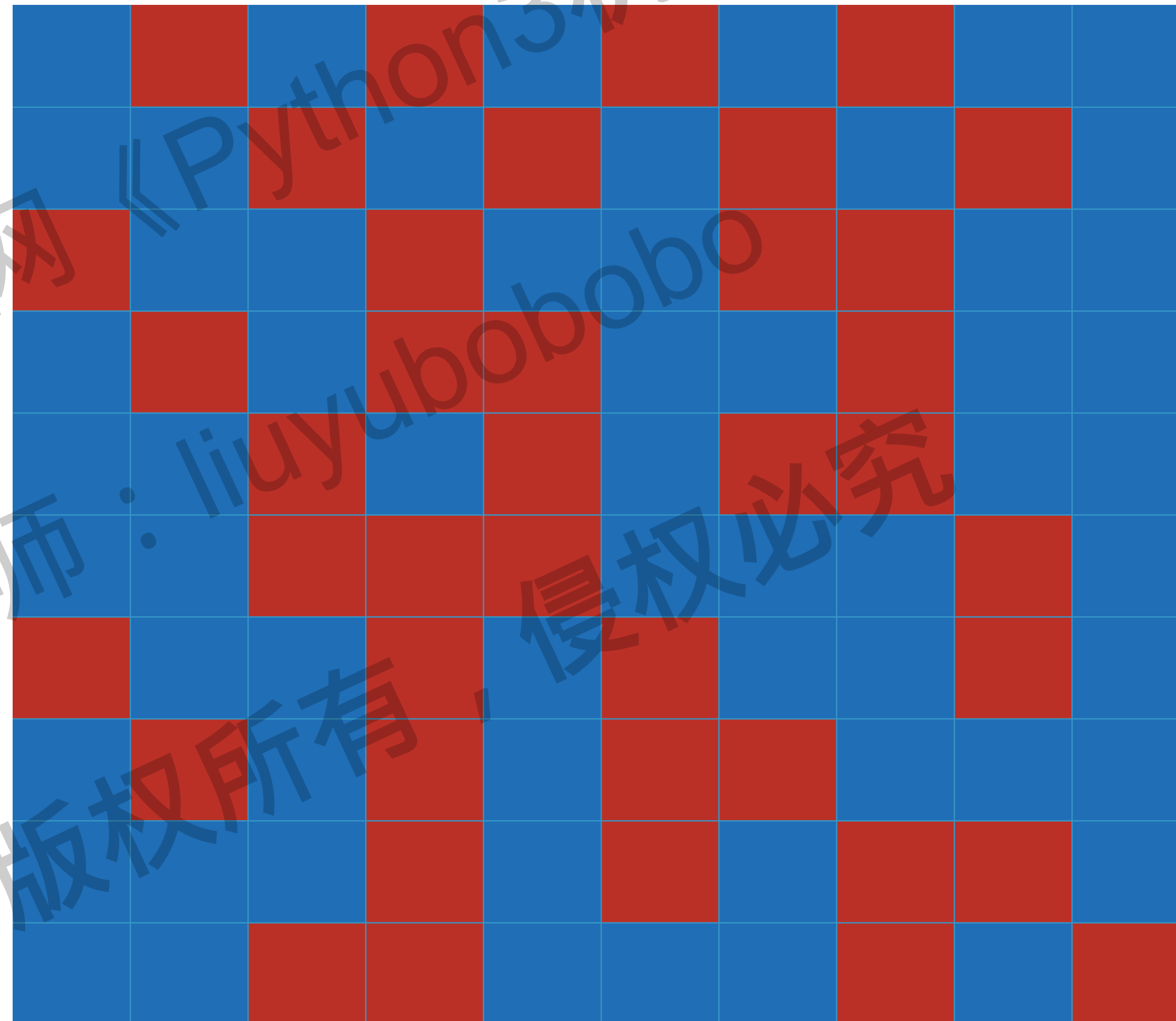
针对特征进行随机采样

Random Subspaces

既针对样本，又针对特征进行随机采样

Random Patches

Random Patches



慕课网《Python3机器学习》

实践：random subspaces和
random patches

讲师：liuyunobobo

版权所有，侵权必究

慕课网《Python3机器学习》

随机森林

讲师：liuyubob0099

版权所有，侵权必究

随机森林

Bagging

Base Estimator: Decision Tree

决策树在节点划分上，在随机的特征子集上寻找最优划分特征

提供额外的随机性，抑制过拟合

慕课网《Python3机器学习》

实践：随机森林

讲师：liuyubobobo

版权所有，侵权必究

随机森林

Bagging

Base Estimator: Decision Tree

决策树在节点划分上，在随机的特征子集上寻找最优划分特征

慕课网《Python3机器学习》

实践：随机森林

讲师：liuyubobobo

版权所有，侵权必究

Extra-Trees

Bagging

Base Estimator: Decision Tree

决策树在节点划分上，使用随机的特征和随机的阈值

提供额外的随机性，抑制过拟合，但增大了bias

更快的训练速度

慕课网《Python3机器学习》

实践：Extra-Trees

讲师：liuyubobobo

版权所有，侵权必究

慕课网《Python3机器学习》

Boosting

讲师：liuyuboboo

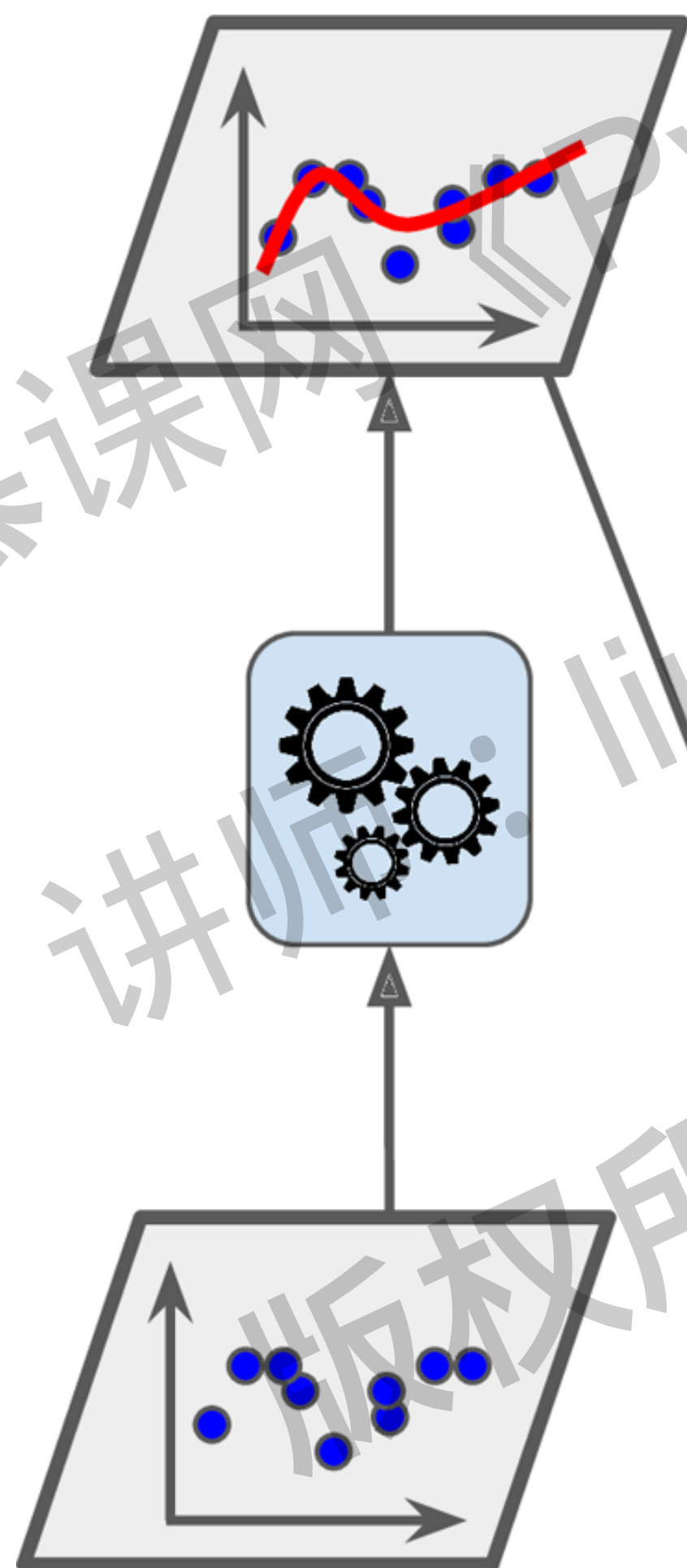
版权所有，侵权必究

Boosting

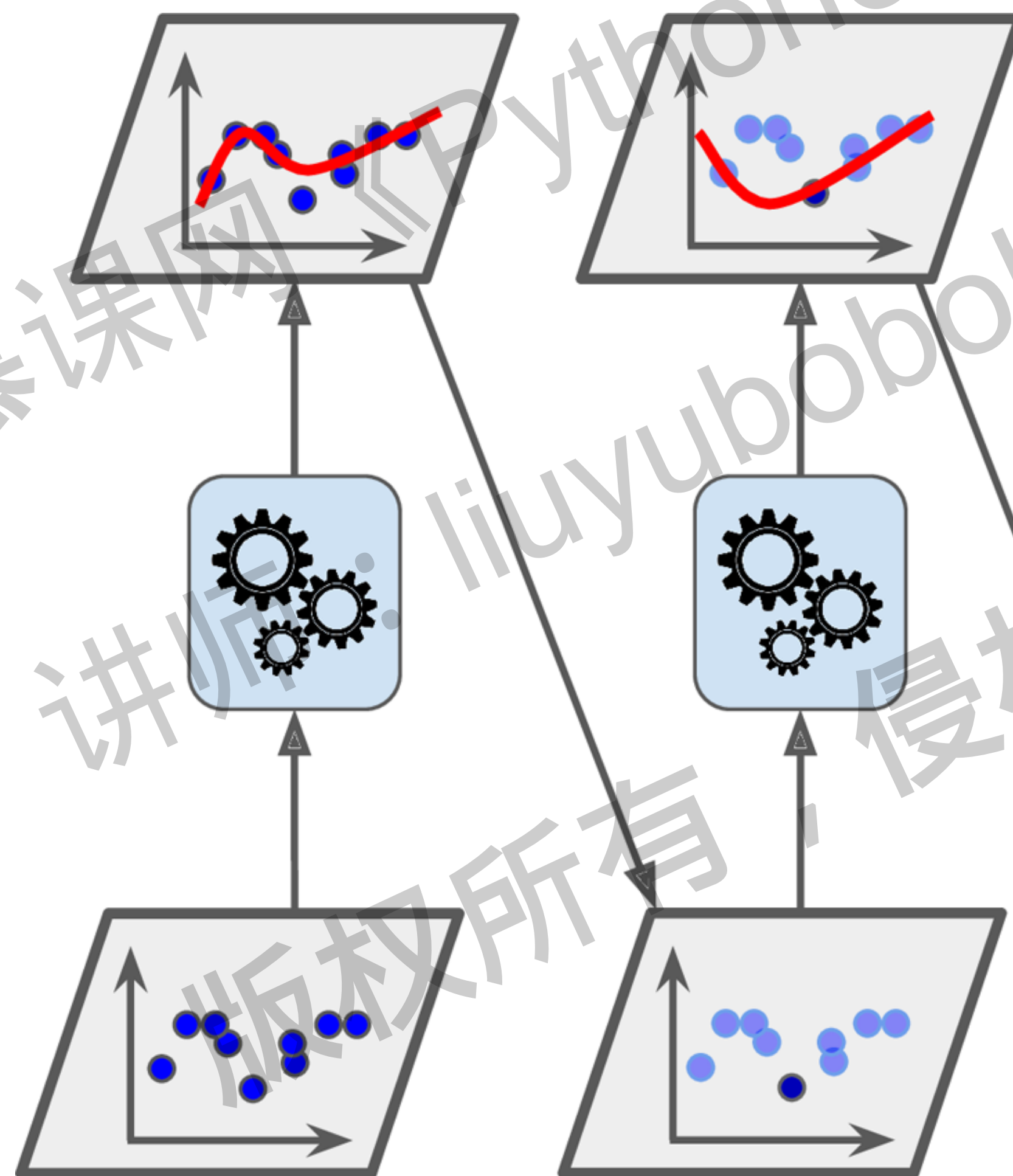
集成多个模型

每个模型都在尝试增强(Boosting)整体的效果

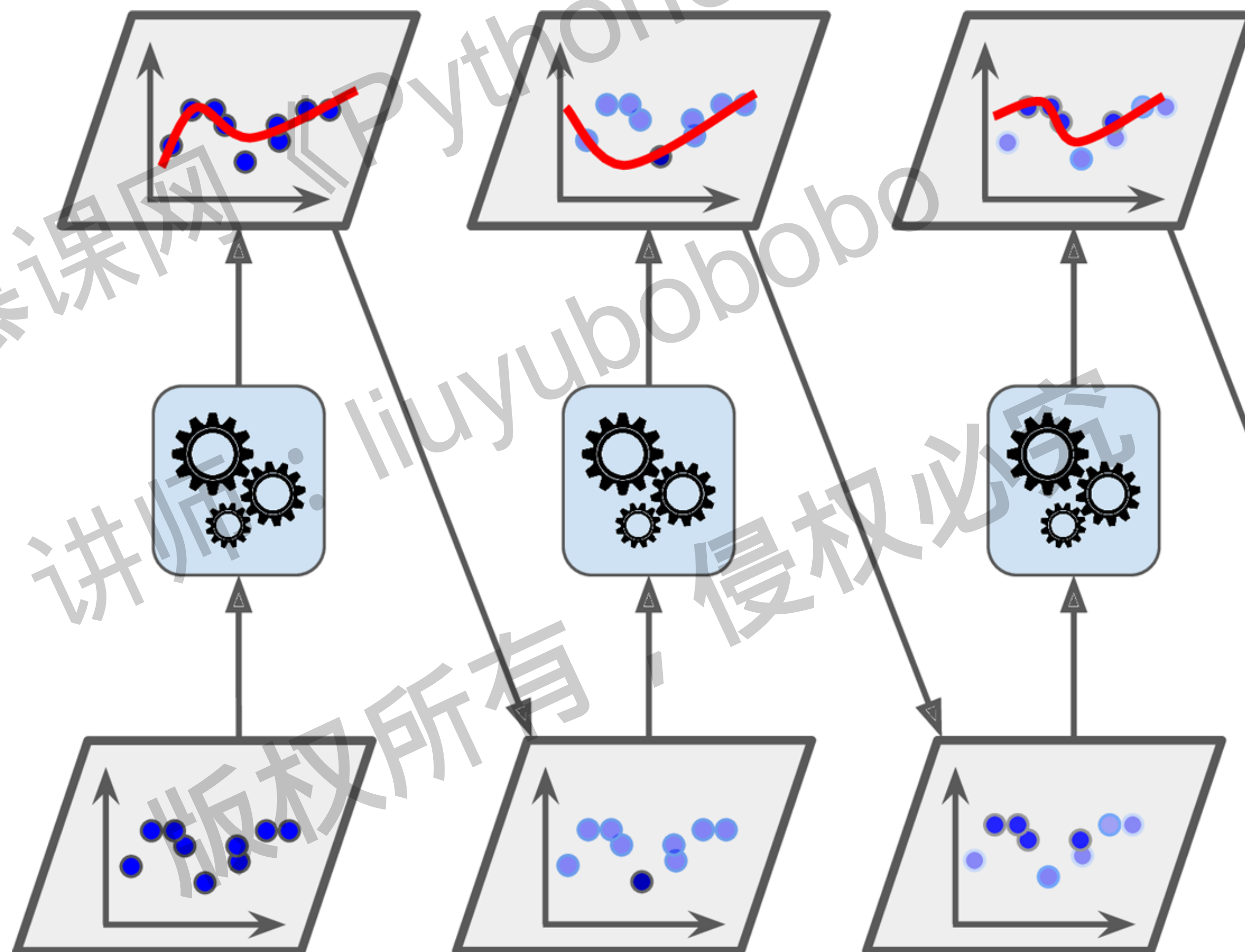
Ada Boosting



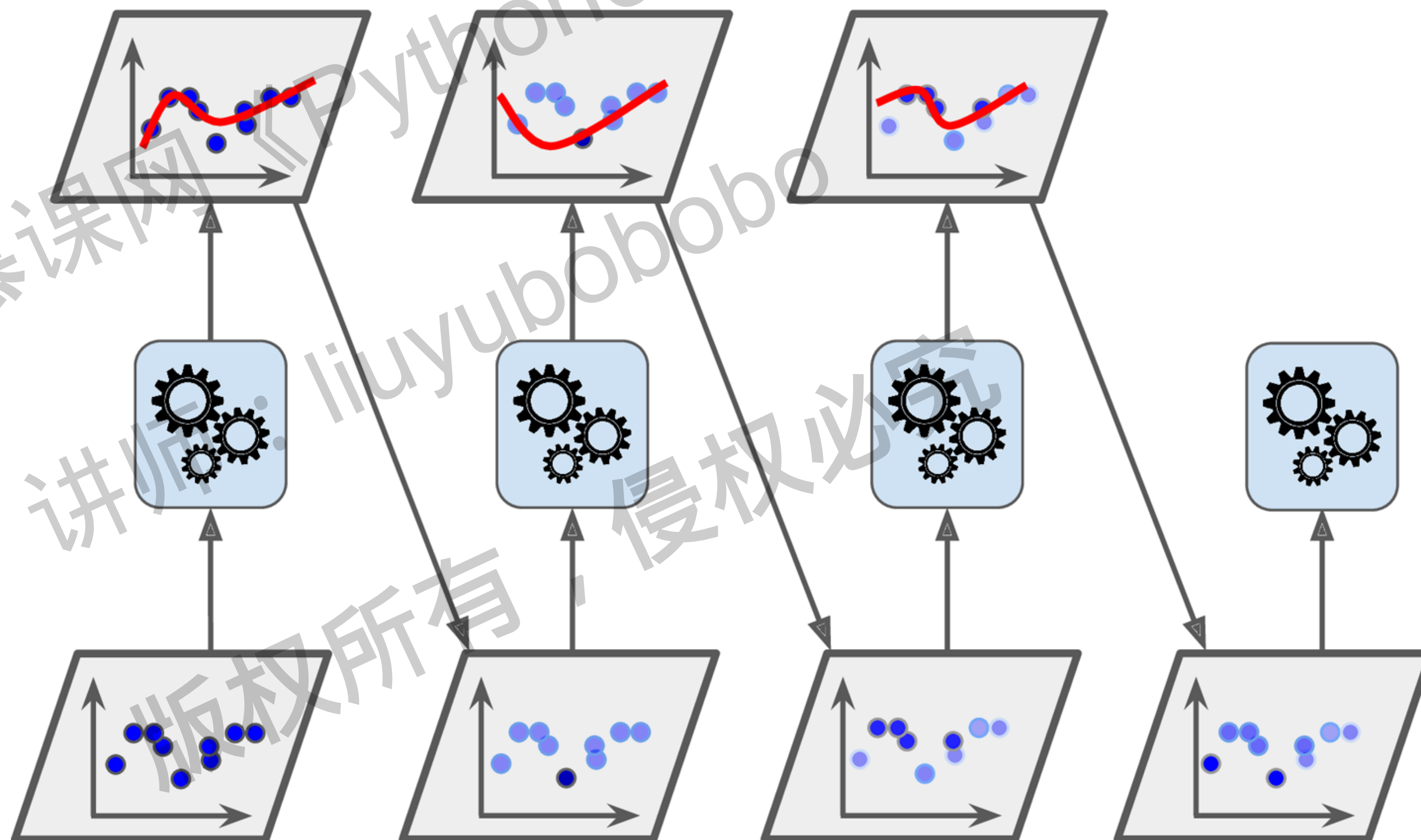
Ada Boosting



Ada Boosting



Ada Boosting



慕课网《Python3机器学习》

实践：Ada Boosting

讲师：liuyuboboo

版权所有，侵权必究

Gradient Boosting

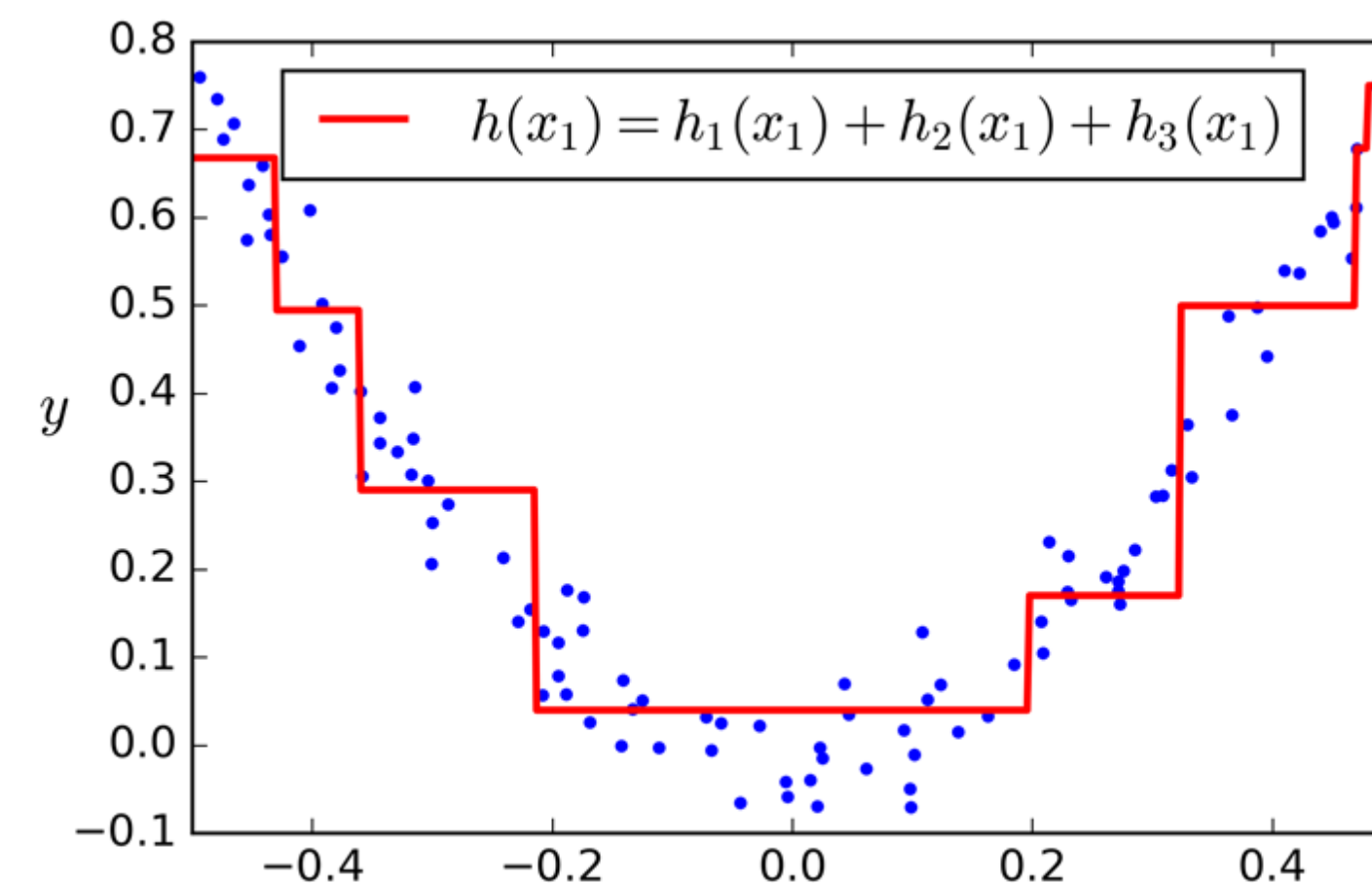
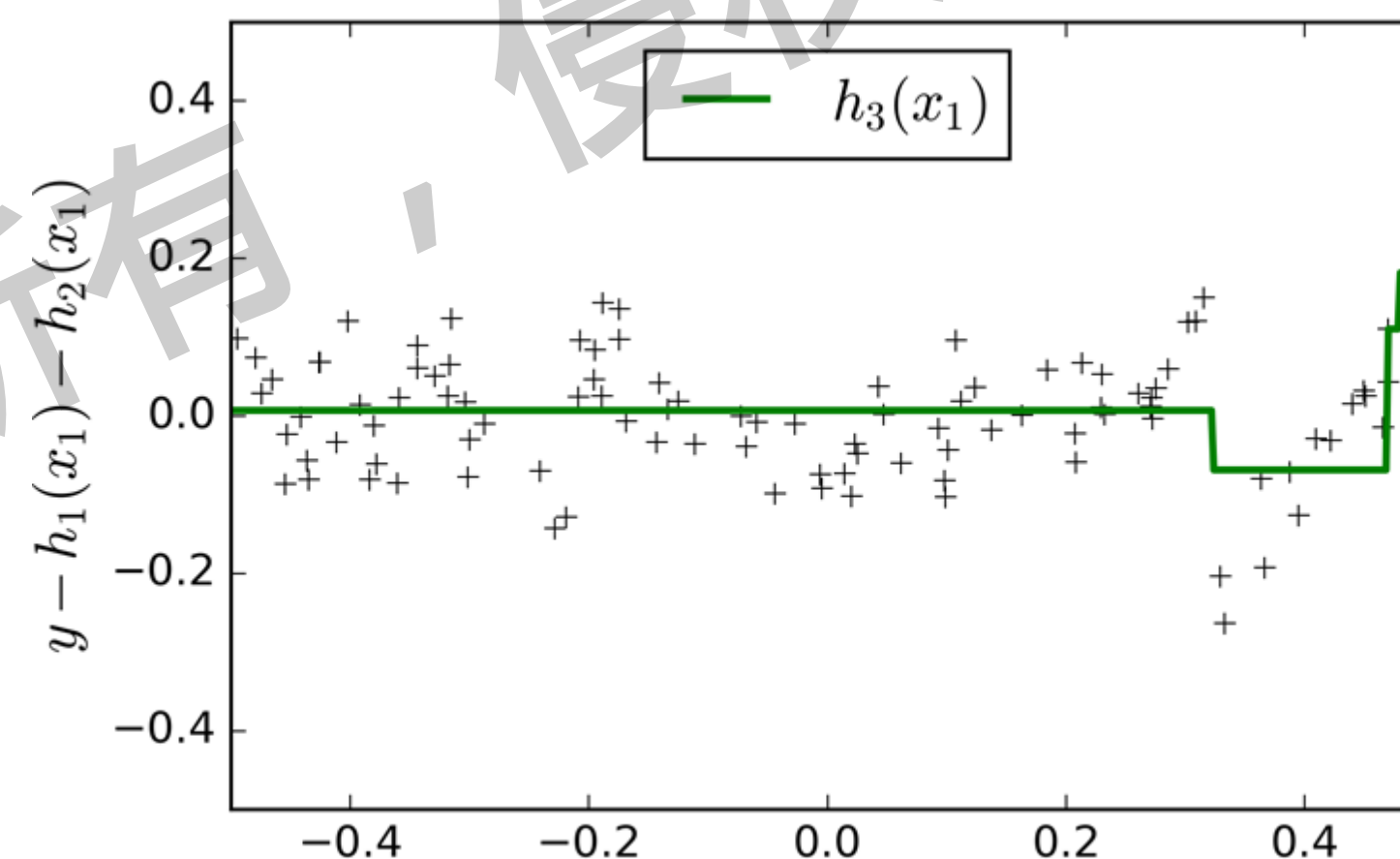
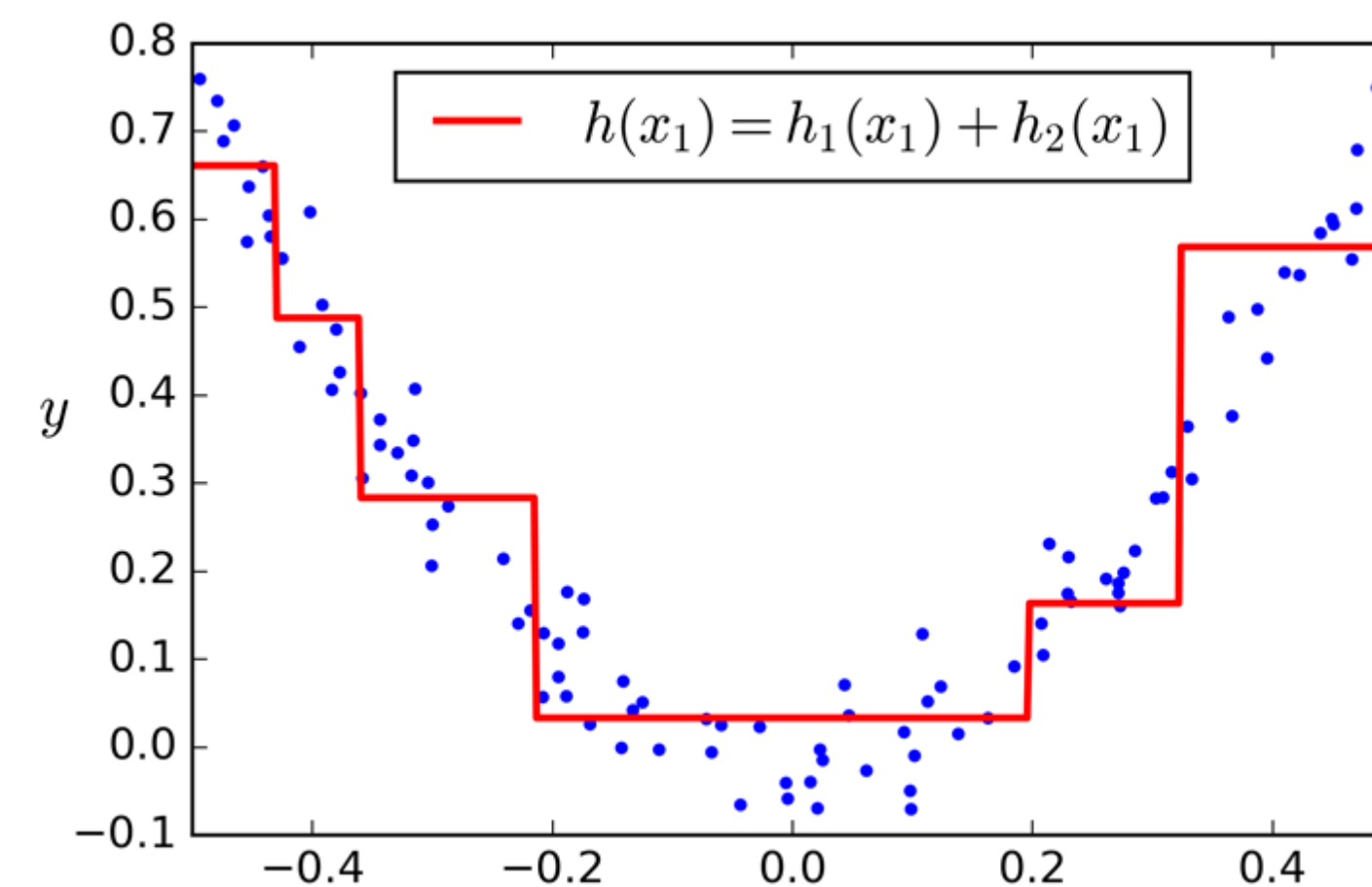
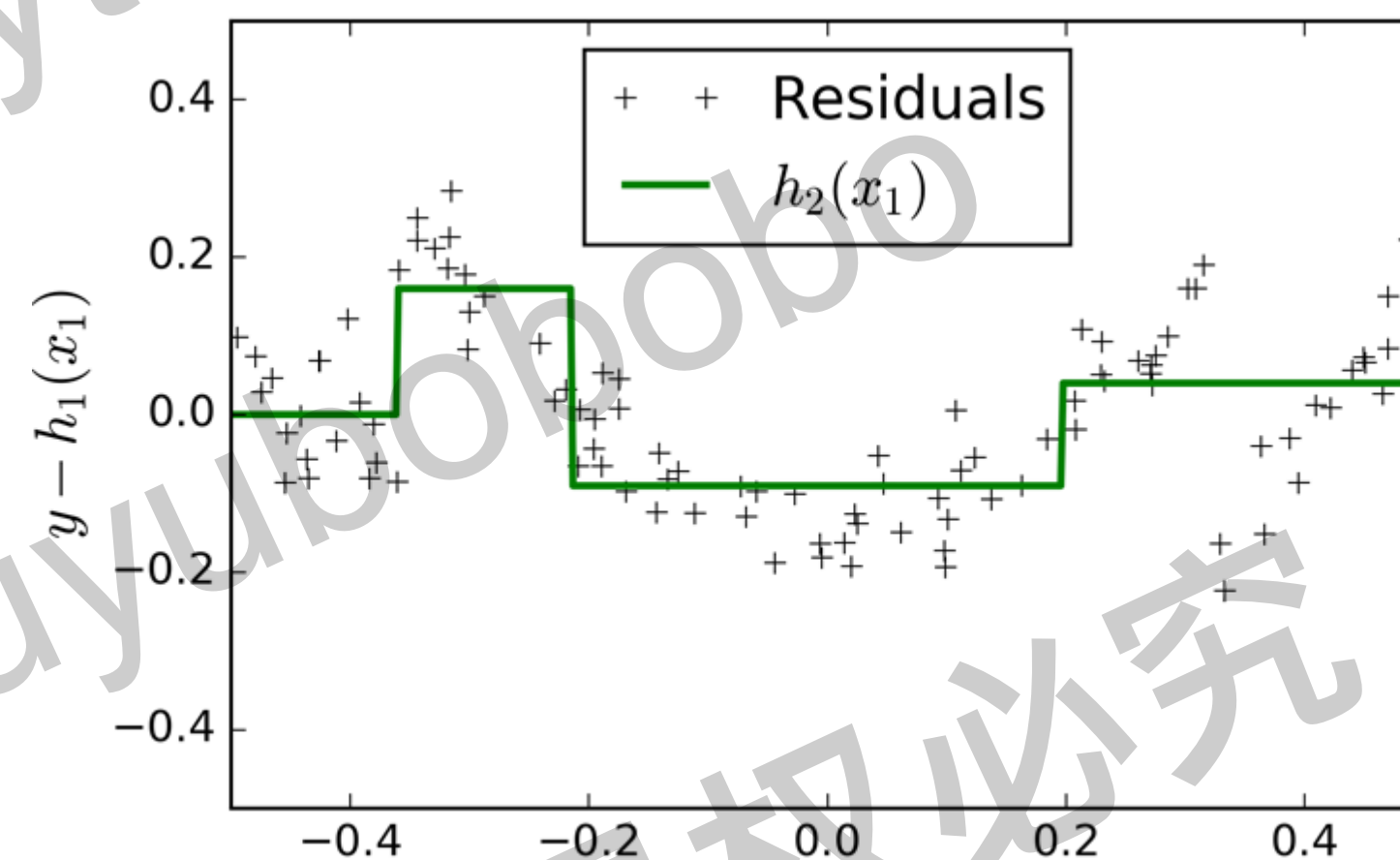
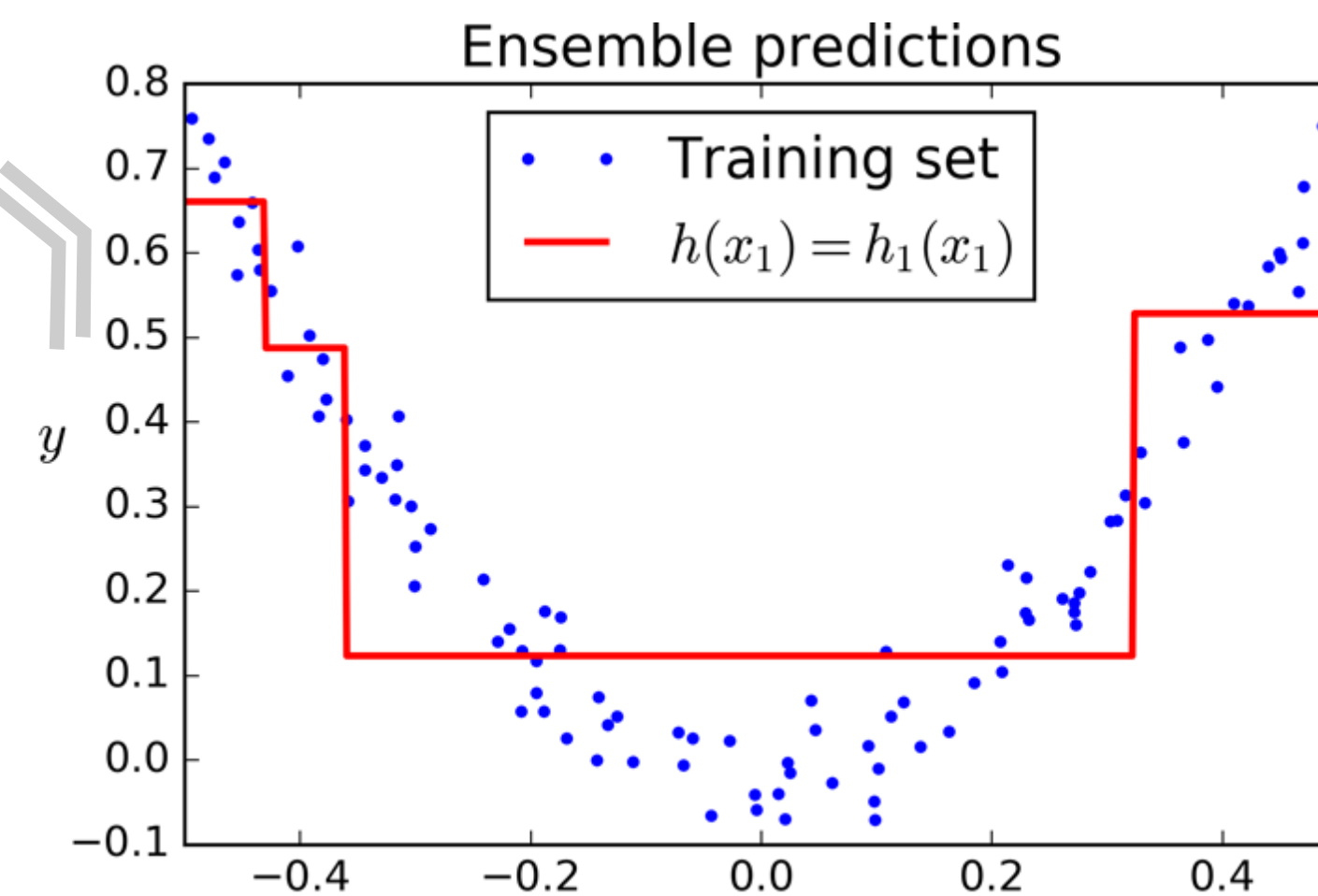
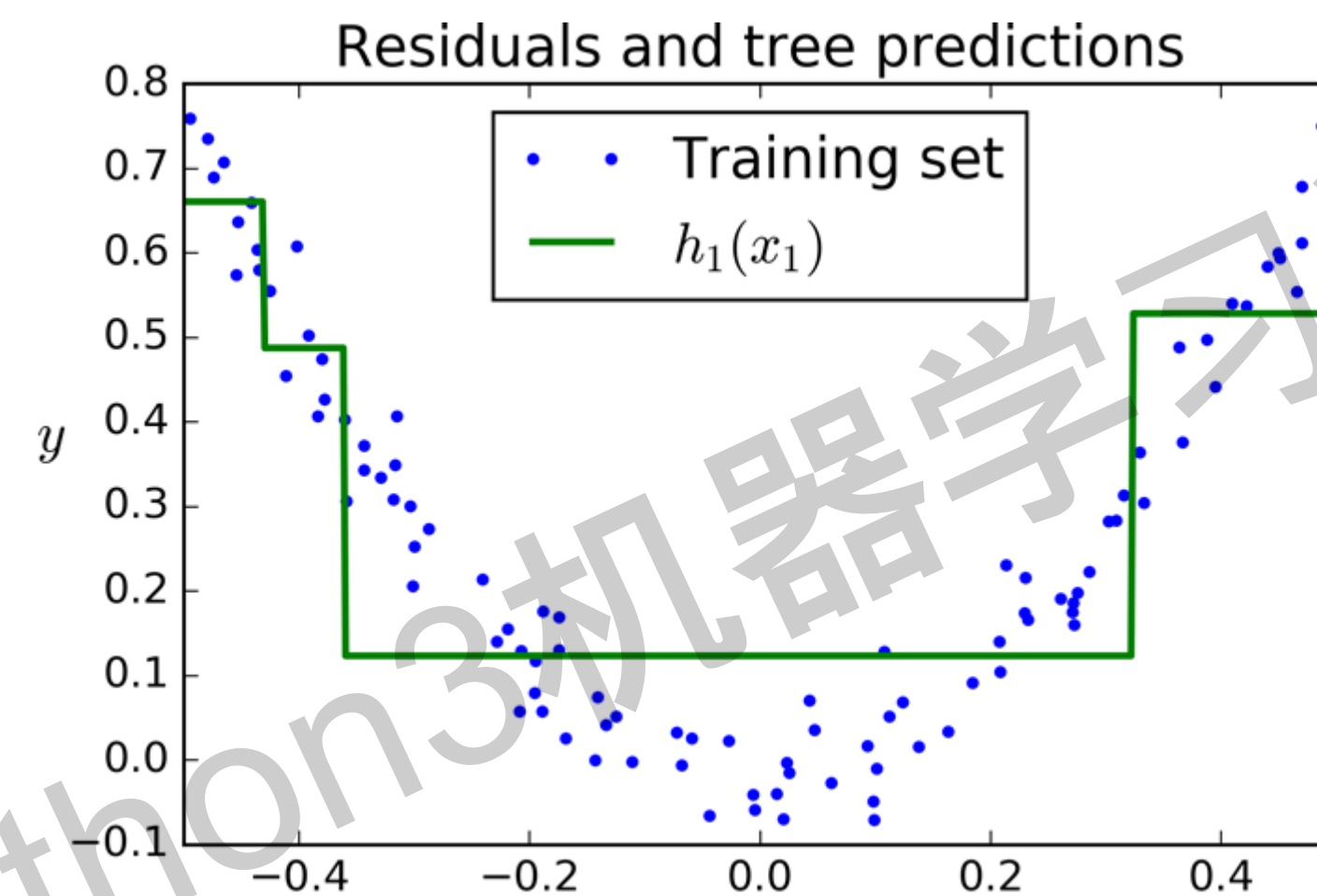
训练一个模型 m_1 , 产生错误 e_1

针对 e_1 训练第二个模型 m_2 , 产生错误 e_2

针对 e_2 训练第三个模型 m_3 , 产生错误 $e_3...$

最终预测结果是: $m_1 + m_2 + m_3 + \dots$

Gradient Boosting



实践：Gradient Boosting

讲师：liuyubobobo
版权所有，侵权必究

慕课网《Python3机器学习》

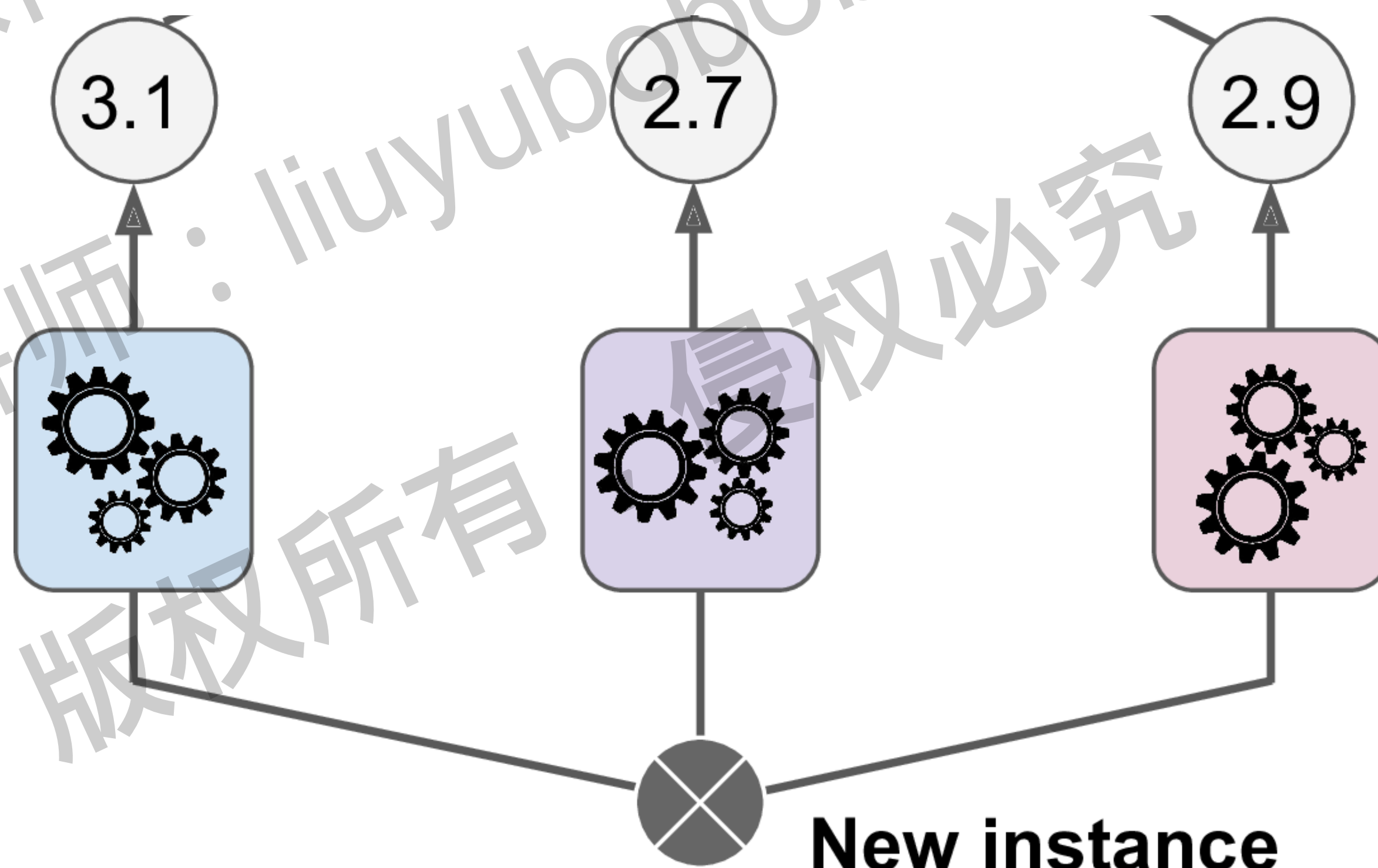
Stacking

讲师：liuyubobobo

版权所有，侵权必究

慕课网《Python3机器学习》

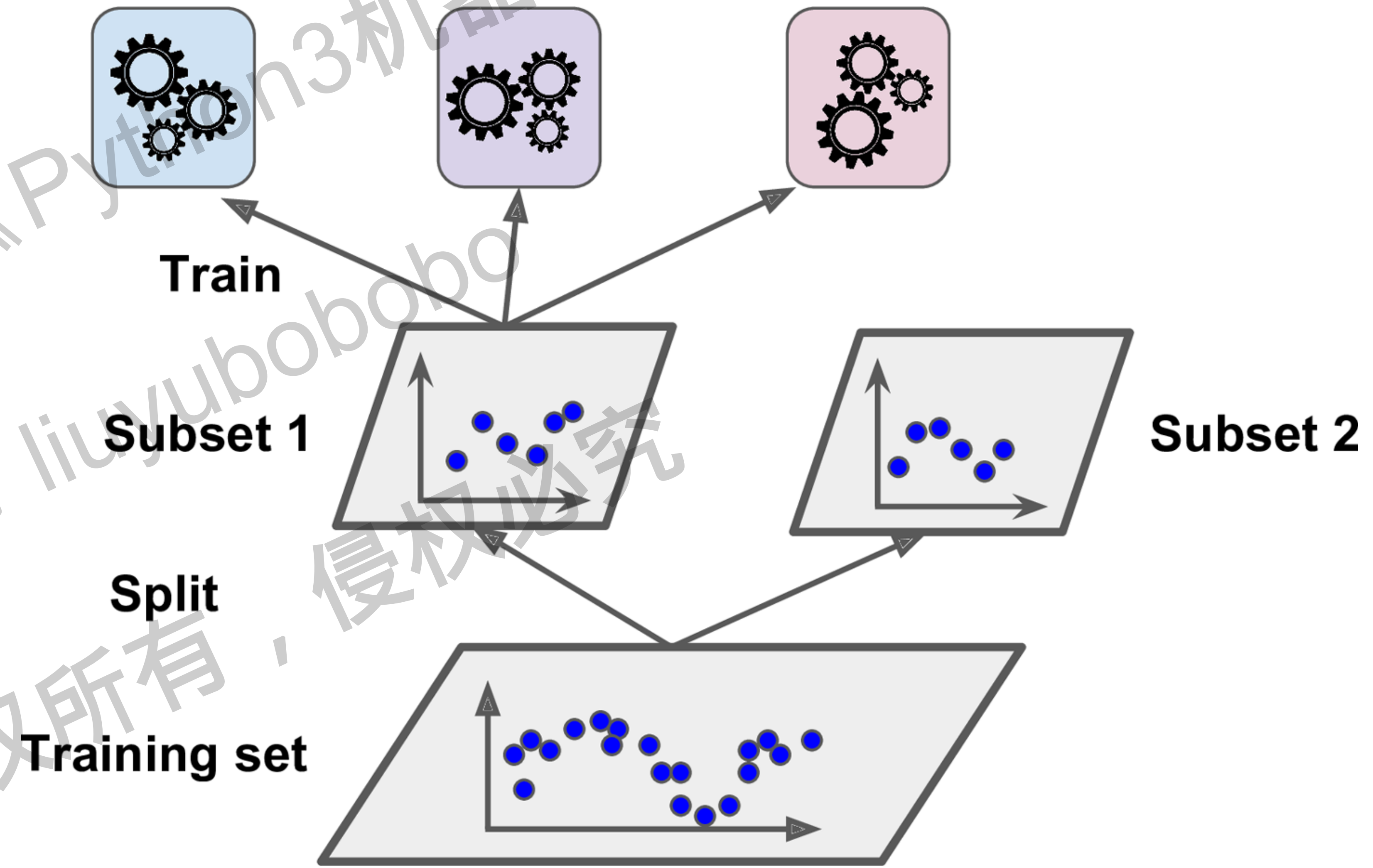
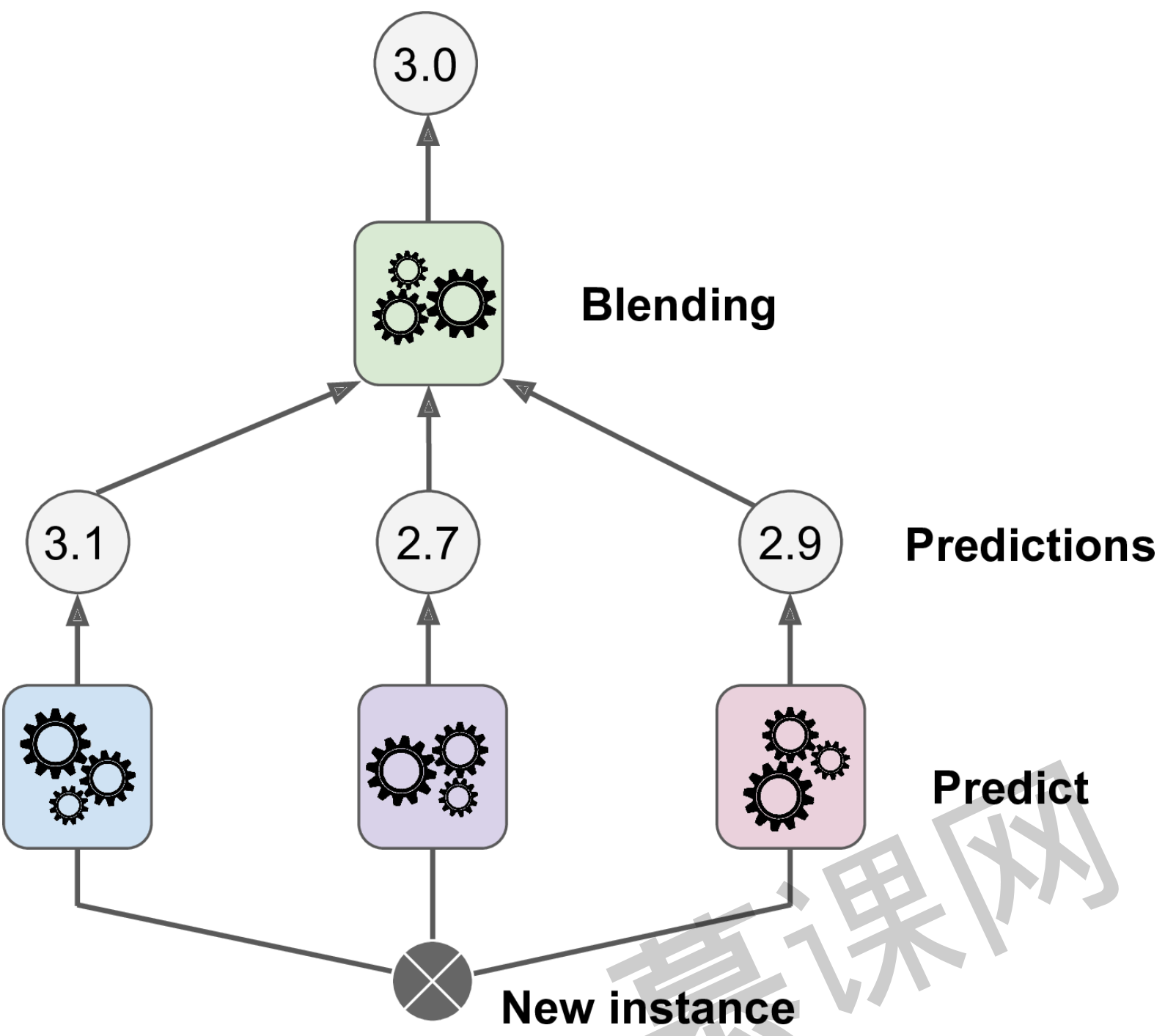
讲师: liuyubobobo
版权所有

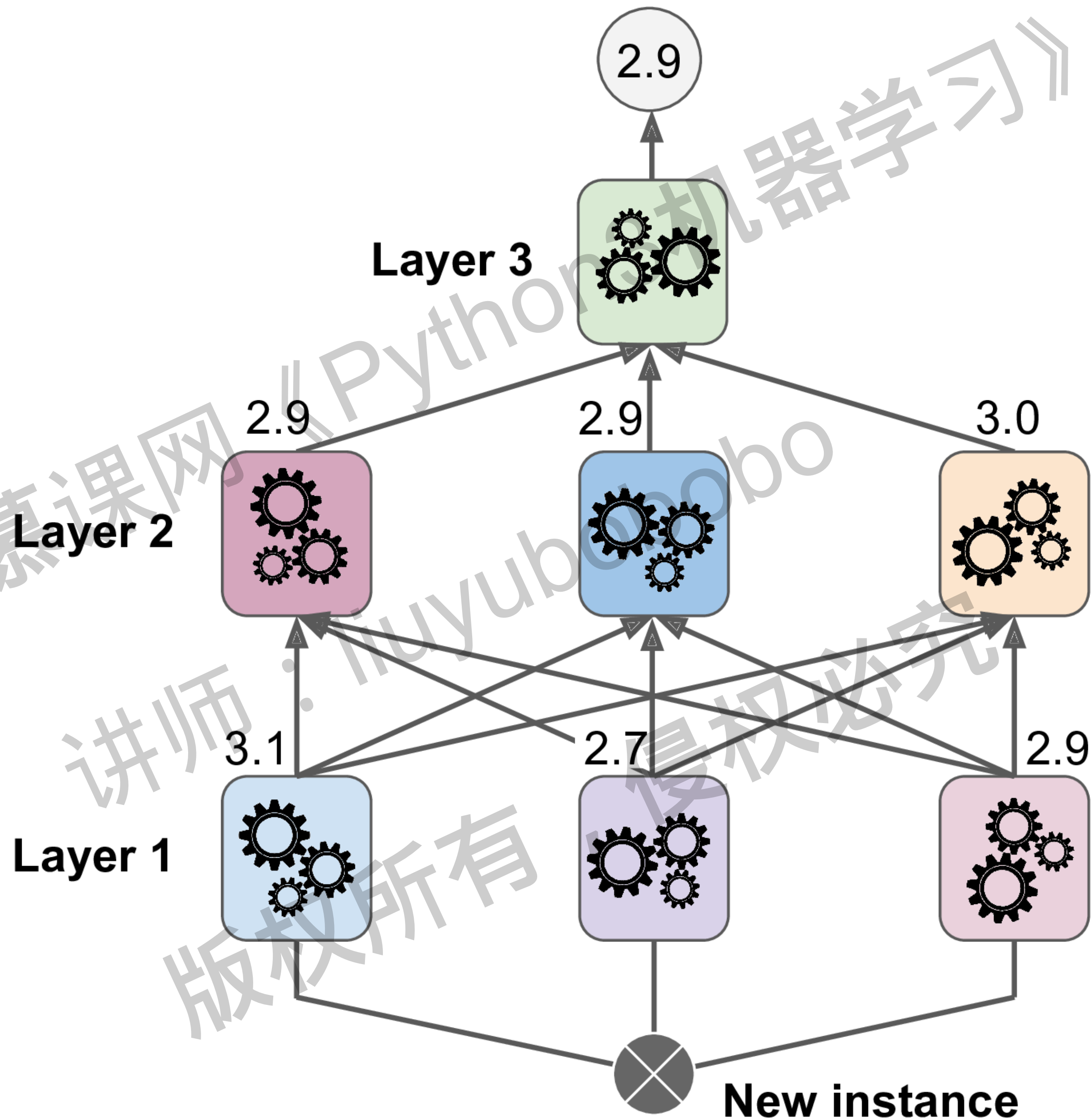


Predictions

Predict

New instance





其他

欢迎大家关注我的个人公众号：是不是很酷



Python 3 玩儿转机器学习

讲师：liuyubobobo

版权所有 侵权必究
liuyubobobo