

《预测未来身高——简单线性回归》

一、教学内容分析

随着时代的飞速发展，技术不断地更新，以人工智能为代表的复杂信息系统早已经和我们的生活息息相关。时代的发展给学生提出了新的要求——了解人工智能，会用人工智能。此外，初中阶段的 python 人工智能内容也为学生进一步学习人工智能知识夯实了基础。本课选取的内容是人工智能领域中较为基础的简单线性回归算法，采用学生自主探究的方式进行教学，旨在帮助学生领会人工智能知识，掌握人工智能应用技能，提升学生借助新技术创造性解决问题的能力，引导学生健康生活，快乐成长。

简单线性回归是人工智能领域的常见模型，它能够反应单个解释变量和连续响应变量之间的关系。简单线性回归的算法主要包括建立目标模型、建立损失模型、更新特征参数等环节。根据学生的认知能力与教材内容的特点，教师引导学生，通过以下活动探究学习：①数据处理（学生对数据进行分析处理，识别数据的主要特征）；②建立目标模型与损失函数模型；③建立模型（分析找寻最优解的过程）；④分析理解简单线性回归算法；⑤使用 python 实现简单线性回归的应用（对人工智能理解以及 python 编程能力的综合运用）。

二、学生特征分析

本课的授课对象是初中一年级的学生，当前他们处在身体发育、思维培养的关键阶段。学生的认知能力有着显著的提升——能够在具体运算的基础上，根据材料进行演绎、抽象和推理，具有一定的系统思维能力，但仍需从具体到形式，从形象到抽象来过渡引导。虽然学生对具体的人工智能的算法所知甚少，但在此前的学习生活中，他们接触了大量相关信息，对人工智能保持着浓厚的兴趣，在之前的学习中也掌握了基础的 python 编程能力，这将有利于建立新的认知结构。学生十分喜爱信息技术课程内容，适应多媒体教学环境。此外，学生的注意力显著提升，精力充沛，个个都是好奇宝宝，学习的内部动机强烈。他们能通过自主探究进行学习。因此，教师创设了探寻自己未来身高的情境，学生通过多个探究活动领会知识，分析解决问题，进而发展自己的计算思维能力。

三、教学目标

基于对《普通高中信息技术课程标准（2017 年版）》、教学内容、学生特征的分析，结合时代背景与学生发展需要，本课的教学目标确定为以下三点：①通过探究、学习，能够理解线性回归算法；②通过思考、分析，能够使用 python，借助 scikit-learn 类库预测出自己的未来身高；③通过探究发现、编程体验等活动，提升计算思维能力。

四、重难点

重点：分析线性回归算法，归纳机器学习的一般步骤

难点：理解梯度下降法

五、教学过程

1. 情境导入：

初中正是学生生长发育的关键时期，学生对自己的未来的身高充满了期待。本课以询问未来身高为起点；抓住学生的注意力，勾起学生对自己未来身高的好奇心，以介绍高尔顿定律为契机（子辈的身高有回到父辈平均身高的趋势，每对夫妇的平均身高 x 与子女的身高 y ，存在着“某种”的对应关系）激发求知欲望。

2. 内容新授

1) 数据处理

学生根据教师提供的学习任务单处理、分析训练集数据，寻找主要特征。

2) 构建回归模型

学生使用学习任务单绘制出已知数据，将数据可视化后，教师引导学生根据自己的探究发现，选择构建自己的回归方程并说明理由。教师适时引导，给予鼓励，帮助学生发现父辈平均身高与子女的身高之间存在着线性对应关系，引出预测函数。

3) 建立损失模型

学生经过环节 1 的自主探究，发现了数据之中所存在的特殊关系。此时，教师提问学生——是否能自己画出一条直线来拟合这些数据？学生比较同桌之间、同学之间的直线是否一致，巧妙制造学生之间的理解冲突。引导学生探讨如何衡量当前直线的合理性，以此探究出使用残差平方和的衡量方法，引出成本函数。

4) 建立训练模型

在学生建立了回归模型与损失模型的基础上，教师引导学生应找到参数的最优解，得到最优的回归模型。进而讲授本课的关键内容，即机器学习的灵魂——建立训练模型，寻找最优解。教师通过“下坡问题”、“数形结合分析”两个例子讲解梯度下降法，并给出相关公式，帮助生理清梯度下降算法的过程及意义。

5) 分析线性回归算法

学生借助学习任务单上的流程图归纳线性回归算法的具体步骤与流程。在学生初步领会算法后，教师以“数据全部用于模型训练还是 80%用于训练, 20%用于测试, 哪种方法效果更好？”制造学生的认知冲突，引出过拟合概念。

6) 编程应用

教师向学生介绍 `scikit-learn` 类库的估计器 API，帮助学生厘清学生编程思路，最终得到自己的未来身高。

3. 拓展延伸

请学生思考: 如果特征空间里加一个维度, 比如省份, 你认为会对回归模型产生什么影响? 学生讨论, 教师最后进行总结, 带领学生使用思维导图, 加深理解简单线性回归算法, 总结归纳机器学习的一般流程。并借此机会学生进行德育教育——在身体生长的关键期间, 应养成良好的生活习惯

六、 辅助资源

附件 1. 学习任务单

附件 1

《预测未来身高》学习任务单

姓名：_____ 班级：_____

一、请根据“父辈与子辈的身高统计表”，根据自己的理解分析样本数据：

编号	父辈平均身高(cm)	子辈身高(cm)
1	1.62	1.70
2	1.65	1.72
3	1.70	1.73
4	1.78	1.80
5	1.83	1.86

父辈与子辈的身高统计表

1. 请你对数据进行简单分析，发现其中的奥秘

图 1.1

2. 我发现数据之间存在着_____（线性/非线性）关系，因此我们选用_____（回归模型/分类模型），这种规律可以采用式子_____表示。请在图 1.1 中尝试表示出来。

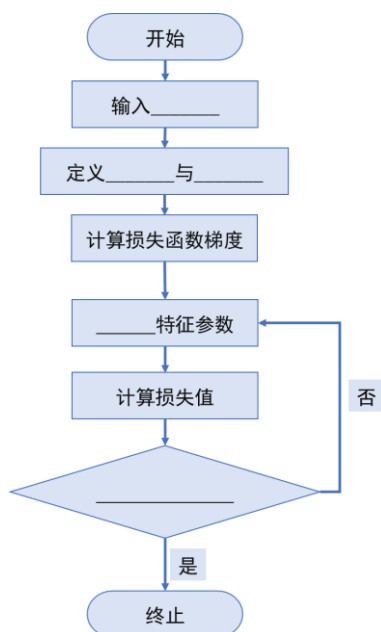
3. 请仔细观察图 1.1 中点与线的关系，你认为我们该如何衡量该模型的拟合效果？

二、万事俱备，只欠东风。目前我们只需通过训练模型找到损失模型的最小值，进而找到回归模型的最优的参数值。

如何找到损失模型的最小值，进而找到最优参数呢？

其实，寻找最优参数（调整参数）的方法有很多，例如：退火算法、蚁群算法等。梯度下降是其中最简单的一种，我们也将重点介绍这一种训练模型。

三、简单线性回归算法分析。



四、运用 scikit-learn 模块完成程序设计。

LinearRegression 会调用 fit 方法来拟合数组 X, y, 并且将线性模型的系数 w 存储在其成员变量 coef_ 中。
 LinearRegression 会调用 predict 方法来对数据进行预测, 输出结果。

五、如果特征空间里加一个维度,比如省份,你认为可能会对回归模型产生什么影响?

答:_____。

六、知识回味。

