Tutorials: NeurIPS - Ariel Data Challenge 2024

Materials

Official Documentations

```
1   Note: Following links direct to local file, you may see kaggle doc
    temporarily.
```

English	中文版
Overview	课题概述
Data	数据介绍

Official Notebooks

```
1   Note: Following links direct to local file, you may see kaggle doc
    temporarily.
```

- [Notebook]Host Starter Solution
- [Notebook]Calibration

Tutorials Documentation

```
1   As this task is considered to be a regression task, you may start with
    following case for better understanding.
```

[Kaggle Notebook]Housing Price Prediction (Linear Regression)

Course Outline

This course is designed to introduce undergraduate students with little machine learning knowledge to the key concepts and approaches necessary to succeed in the NeurIPS - Ariel Data Challenge 2024. The course will consist of several one-hour lectures, each focusing on a specific aspect of the challenge.

Lecture 0: Preliminary Knowledge

- **Python Basics**: Introduction to Python programming, focusing on the necessary skills for data analysis.
- **Kaggle and Notebook Environments**: Introduction to Kaggle platform, competition basics, and using Kaggle Notebooks for coding.
- **Essential Machine Learning Concepts**: Basic concepts of machine learning, including training, testing, features, and labels.

▶ 【大纲】讲座0：预备知识

```
1  - **Python 基础**：Python 编程的介绍，重点讲解数据分析所需的技能。
2  - **Kaggle 平台和 Notebook 环境**：Kaggle 平台介绍，竞赛基础，以及使用 Kaggle
   Notebooks 进行编程。
3  - **机器学习基础概念**：机器学习的基本概念，包括训练、测试、特征和标签。
```

Lecture 1: Introduction to Ariel Data Challenge 2024

- **Background (5 minutes)**: Introduction to exoplanets, their importance, and the role of the Ariel Mission.
- **Research Objective (5 minutes)**: Overview of the challenge - extracting faint exoplanetary signals from noisy data.
- **Data Overview (10 minutes)**: Types of data used in the competition - image data (2D spectral images), sequence data (time-series), and Ariel data.
- **Evaluation Metrics (10 minutes)**: Gaussian Log-Likelihood (GLL) function and how submissions are evaluated.
- **Introduction to Regression/Prediction Task (20 minutes)**: Explanation of regression tasks and how prediction models are used to solve the challenge.
- **Quick Start on Kaggle for Submission (20 minutes)**: Setting up a Kaggle account, accessing the competition, navigating Kaggle, and making a submission.

▶ 【大纲】讲座1：Ariel 数据挑战赛 2024 介绍

```
1  - **背景介绍（5 分钟）**：系外行星的介绍及其重要性，以及 Ariel 任务的作用。
2  - **研究目标（5 分钟）**：挑战赛概述 - 从噪声数据中提取微弱的系外行星信号。
3  - **数据概述（10 分钟）**：竞赛中使用的数据类型 - 图像数据（二维光谱图像）、序列数据（时间
   序列）、Ariel 数据。
4  - **评估指标（10 分钟）**：高斯对数似然（GLL）函数以及如何评估提交结果。
5  - **回归/预测任务介绍（25 分钟）**：解释回归任务以及如何使用预测模型解决挑战。
6  - **Kaggle 快速开始与提交（5 分钟）**：设置 Kaggle 账户，访问竞赛，导航 Kaggle 界面，
   提交结果。
```

Lecture 2: Official Solution Walkthrough

Focus: Reviewing the official solution notebook provided in the competition.

- **Notebook Walkthrough (~1 hour)**: Step-by-step explanation of the code in the official solution.
- **Q&A**

▶ 【大纲】讲座2：官方解决方案讲解

```
1  **重点**：回顾竞赛中提供的官方解决方案。
2
3  - **notebook 讲解（~1 小时）**：逐步解释官方解决方案中的代码.
4  - **答疑**
```

Lecture 3: Promising Solutions to SOTA

▶ 【大纲】讲座3：可能有效的显著提升模型表现的技术

```
1   - **技术概述**：从数据形式和任务两个角度出发，提升模型表现.
2   - **答疑**
```

【大纲】讲座3：可能有效的显著提升模型表现的技术

```
1   - **技术概述**：从数据形式和任务两个角度出发，提升模型表现.
2   - **答疑**
```