作业 2：线性模型、决策树和强化学习

清华大学软件学院

人工智能导论，2025 年春季学期

介绍

本次作业需要提交说明文档（PDF 形式）和 Python 的源代码。注意事项如下：

- 本次作业满分为 100 分，附加题 5 分，得分超过 100 分按 100 分记。
- 除简答题、编程题外的题目，请给出必要的解答过程，只有答案且过于简略的回答会酌情扣除一定分数。题目要求直接回答或只汇报结果的题目，则不需要给出过程或分析。
- **请不要使用他人的作业，也不要向他人公开自己的作业，复制网上内容须在报告中说明**，否则将受到严厉处罚，作业分数扣至-100（即倒扣本次作业的全部分值）。
- 完成作业过程中，如果使用了大模型辅助（如润色文笔、询问知识点等），请在作业末尾声明使用的方式和程度（不影响作业评分）。**禁止直接粘贴大模型输出的文本**，否则会扣除一定的作业分数。
- 统一文件的命名：`{学号}_{姓名}_hw2.zip`。**所有解答和实验报告请写在一个 pdf 文件中**，和代码一起压缩上传。

1 简答题（25 分）

1. 什么样的场景适合使用 K 折交叉验证（K-fold Cross Validation）？参数 K 越大越好吗？
2. 线性模型中为什么要引入基函数？基函数越复杂，学习的效果一定会越好吗？
3. 随机森林使用哪些方法增加单棵决策树的多样性？
4. 为什么说相比使用蒙特卡洛（Monte-Carlo）采样的策略梯度（Policy Gradient）方法，Actor-critic 方法的方差更小？
5. 为什么使用神经网络（Neural Network）的 Q-Learning 不能保证收敛到最优状态动作值函数 Q^*？使用基于策略的方法（Policy-based）方法能保证收敛到全局最优策略吗？

2 ID3 算法的次优性（25 分）

考虑以下训练数据，其中 $\mathcal{X} = \{0,1\}^3, \mathcal{Y} = \{0,1\}$:

$$(x_1, y_1) = ((1,1,1),1)$$
$$(x_2, y_2) = ((1,0,0),1)$$
$$(x_3, y_3) = ((1,1,0),0)$$
$$(x_4, y_4) = ((0,0,1),0)$$

1. 写出基于信息增益准则的 ID3 算法构造最大深度不超过 2 的决策树的过程，包括信息增益的计算过程（当两个属性信息增益相同时，任选其一进行分裂）；并说明训练误差将至少为 1/4（即至少有一个训练样本分类错误）。
2. 写出一棵深度为 2 的决策树，其训练误差为 0。

3 线性模型与 AlphaGoZero（50pt+5pt）

本题为编程题，代码和相关文档在下发文件的`./code`目录下。

问题背景 在第一次作业中，我们实现了基于 UCB 公式的 MCTS 搜索，它在比较简单的问题上有不错的表现。但在围棋中，由于局面状态数极大，使用蒙特卡洛采样和 UCB 公式的 MCTS 无法取得很好的表现。为了解决这个问题，AlphaZero 中，引入了策略（Policy）和价值（Value）网络和 PUCB 公式用来修正决策和代替 rollout 采样。本次作业，我们将实现一个简单的 AlphaGoZero 训练流程。

我们将使用线性模型作为策略和价值模型，其中策略模型参数为 $W_\pi \in R^{O\times A}, \boldsymbol{b}_\pi \in R^{1\times A}$，价值模型参数为 $\boldsymbol{w}_v \in R^{O\times 1}, \boldsymbol{b}_v \in R^1$，其中 O 输入特征的维度，A 是动作空间的大小。推理时，对于输入的一个批次的 B 条局面的观测值 $X \in R^{B\times O}$，我们记 $P \in R^{B\times A}$ 和 $\hat{P} \in R^{B\times A}$ 分别是 B 条数据的 MCTS 策略和策略模型输出策略，$\boldsymbol{z} \in R^{B\times A}$ 和 $\boldsymbol{v} \in R^{B\times A}$ 分别是 B 条数据 MCTS 计算的价值和价值模型预测的价值。策略和价值模型的输出计算方式为:

$$\hat{P} = \mathrm{softmax}(XW_\pi + \mathbf{1}_B \boldsymbol{b}_\pi)$$
$$\boldsymbol{v} = \tanh\left(X\boldsymbol{w}_v + b_v \mathbf{1}_B\right)$$

其中，$\mathbf{1}_B$ 是 B 行的全 1 向量。而 AlphaZero 中，使用的损失函数为:

$$L = L_{\text{policy}} + L_{\text{value}} + c\|\theta\|^2 = (z-v)^2 - \pi^T \log \hat{\pi} + c\|\theta\|^2$$

对于一组 B 条训练样本，忽略正则项，损失函数可以写成:

$$L = L_{\text{policy}} + L_{\text{value}} = \frac{1}{B}[(\boldsymbol{z}-\boldsymbol{v})^T(\boldsymbol{z}-\boldsymbol{v}) - \mathrm{tr}\left(P^T \log \hat{P}\right)]$$

我们可以计算出，使用我们的线性模型，策略损失函数对模型参数的梯度为:

$$\nabla_{W_\pi} L_{\text{policy}} = X^T(\hat{P} - P)/B$$
$$\nabla_{\boldsymbol{b}_\pi} L_{\text{policy}} = \mathbf{1}_B{}^T(\hat{P} - P)/B$$

任务目标 在本题中，你需要补全代码框架中线性模型训练、使用 PUCB 的 MCTS 和训练数据收集相关的代码。请完成以下内容，根据要求使用和修改 `code` 路径下的代码，提交你的代码和实验报告。**提交时请删除 `*.so`、`*.pyd` 和 `*/build/`等临时文件，仅提交代码，本题的文字报告请和其他题目写在同一个文档中提交。**

1. 计算对于一组 B 条训练样本，价值模型的损失函数 L_{value} 对参数 $\boldsymbol{w}_v, b_v$ 的梯度（写出计算过程）。
2. 参考策略模型的相关实现，补全 `model/linear_model.py` 中空缺的价值模型损失函数和梯度计算，用 `check_grad` 函数验证梯度计算是否正确。
3. 补全 `mcts/puct_mcts.py` 中空缺的内容，实现使用 PUCB 的 MCTS 算法。
4. 补全 `alphazero.py` 中空缺的内容，并选取适当的参数，进行 AlphaGoZero 训练，绘制训练过程中对 Random Player 的胜率/不输率变化的折线图，并汇报你选取的参数。
5. 改变 C 或 n_search 的值，同样绘制对 Random Player 的胜率/不输率变化的折线图，分析该参数是如何影响模型训练和最终胜率的。（二选一完成即可）
6. **[附加题]（3pt）**给线性模型增加一个基函数，分析你选取的基函数对于模型学习棋盘局面的策略和价值有什么好处，并通过实验验证其是否有效。

提示：

1. **动手之前，请仔细阅读 `README.md`**，并仔细阅读代码中相关的注释文本。请尽量不要修改代码中未要求修改的部分，如确有必要请提交前与助教沟通。
2. 你可以复用上一次作业中 MCTS 的代码。
3. 使用 `alphazero.py` 训练并确认代码实现正确后，你可以使用 `alphazero_parallel.py` 进行高效的并行训练。
4. 以下是一些可能有帮助的矩阵求导公式（注意矩阵运算没有交换律）：
   - $\frac{\partial \boldsymbol{g}(\boldsymbol{v}(\boldsymbol{x}))}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{v}(\boldsymbol{x})}{\boldsymbol{x}} \frac{\partial \boldsymbol{g}(\boldsymbol{v})}{\boldsymbol{v}}$，其中 $\boldsymbol{x}$ 是向量，$\boldsymbol{v}, \boldsymbol{g}$ 是向量值函数；
   - $\frac{\partial \boldsymbol{x}^T \boldsymbol{x}}{\partial \boldsymbol{x}} = 2\boldsymbol{x}$，其中 $\boldsymbol{x}$ 是向量；
   - $\frac{d}{dx}\tanh(x) = 1 - \tanh^2(x)$；
   - 对函数 $g : R \to R$ 和向量 $\boldsymbol{x} = (x_1, x_2, ...)^T$，记 $g(\boldsymbol{x}) = (g(x_1), g(x_2), ...)^T$，则 $\frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}} = \text{diag}(g'(\boldsymbol{x}))$。
5. **推荐参考材料**：课件，以及 *A Survey of Monte Carlo Tree Search Methods*[1]和 AlphaZero 论文的补充材料[2]。

4 提交格式

- 请先删除 `*.so` `*.pyd` 等文件和环境目录下的 `build` 文件夹，再将你的代码目录内**所有代码文件**和你的**文字报告**打包提交。统一文件的命名：`{学号}_{姓名}_hw2.zip`。
- 请将本次作业所有问题回答写在同一份报告中，报告请导出为 **pdf 格式**。

[1] `https://repository.essex.ac.uk/4117/1/MCTS-Survey.pdf`

[2] `https://www.science.org/doi/suppl/10.1126/science.aar6404/suppl_file/aar6404-silver-sm.pdf`