项目分析

### **1. 基本交易策略设计**

你描述了一个**日内交易策略**，基于价格上下限来设置止损和止盈。具体内容如下：

* **投资时间范围（Investment Horizon）**：  
  + 每天开盘时买入，在收盘时退出，不持有隔夜头寸。
* **开盘价 (Open Price) x**：  
  + 这是你进行交易决策的基础。当天的开盘价是用于计算止损和止盈的重要参考点。
* **止损限制 (Stop Loss Limit, alpha)**：  
  + 如果当天的最低价 (low price) 小于 open \* (1 - alpha)，那么触发止损。
  + 在这种情况下，损益 (P&L) = open \* alpha，表示你在下跌到一定程度时退出，避免进一步的亏损。
* **止盈限制 (Take Profit Limit, beta)**：  
  + 如果当天的最高价 (high price) 大于 open \* (1 + beta)，那么触发止盈。
  + 在这种情况下，损益 (P&L) = open \* beta，表示你在赚到一定比例时退出，锁定收益。
* **超参数调整**：  
  + **alpha 和 beta** 都是**超参数（hyperparameters）**，可以在训练和回测中调整，以找到最优组合来最大化策略的表现。

### **2. 信号生成与模型训练**

* **信号生成**：  
  + 你基于一个选定的 alpha 和 beta 组合，可以根据当天的开盘价、最高价、最低价和收盘价来生成一个**买入信号**。
* **训练数据的构建**：  
  + 利用过去的交易数据，你可以根据不同的 alpha 和 beta 生成训练数据集。
  + 训练集中的每一行可以包含市场数据（如开盘价、最高价、最低价、收盘价等）以及生成的买入信号（1 代表买入，0 代表不买入）。
* **模型训练和验证**：  
  + 利用这些数据，可以训练一个机器学习模型来预测买入信号。
  + 训练好模型后，可以在验证集上检验它的表现，并计算每种预测结果的收益。

### **3. 目标与模型选择**

* **目标**：
  + 你的目标是找到使得总 P&L 最大的模型。因此，模型的选择可以基于预测的 P&L 总和。
* **关于 Confusion Matrix**：
  + **Confusion Matrix** 是用来评估分类模型性能的工具，主要用于衡量预测信号的准确性（例如买入信号 vs. 实际结果）。
  + 在这种策略中，如果你有兴趣了解你的模型在预测上涨和下跌时的表现，可以使用 Confusion Matrix。它可以帮你量化模型在不同市场情况下的准确性，但它并不是直接用来优化收益的。
  + 换句话说，Confusion Matrix 可以帮助你了解模型是否准确地产生买入信号，但最终目标还是最大化收益。

### **4. 集成策略（Ensemble Method）**

* **多个模型形成集成策略**：
  + 你可以使用不同的模型（比如 CNN, RNN, KNN, Logistic Regression）来生成买入信号。
  + 这些模型可能对市场特征的捕捉有不同的表现，因此你可以为它们分配不同的权重，最终形成一个集成策略。
  + 集成策略通过结合多个模型的预测结果，可以增强整个策略的鲁棒性。

### **5. 教授回复的分析**

* **系统在第一阶段应给出“买入”信号以及获胜的概率**：  
  + 在第一阶段，系统需要给出买入信号，并最好附带一个获胜的概率，这个概率可以看作是模型对信号的置信度。
  + 这个“获胜概率”通常可以通过机器学习模型的**softmax 输出**获得。Softmax 会输出一个值，表示模型对某种信号的信心，这个可以理解为买入成功的概率。
* **交易规模是概率的函数**：  
  + 教授提到交易的规模可以是概率的函数，这意味着如果模型对某次交易的信心很高（即获胜概率很高），可以加大交易的规模（例如购买更多的份额）；反之，如果信心较低，则可以减少交易规模或避免交易。
  + 在目前的阶段，你可以假设每次只买卖一股来简化计算，但在以后可能要考虑动态调整交易量以提高收益。
* **止损与止盈的设置**：  
  + 教授提到你可以自由选择初始的投资金额，但必须设置一个止损（x%）和止盈（y%）的限制。这个也是为了风险控制，确保在大幅下跌或盈利时及时退出。

### **6. 接下来的步骤建议**

1. **超参数调整**：  
   * 尝试不同的 alpha 和 beta 组合。
2. **模型训练**：  
   * 构建训练数据集，基于不同市场条件训练多种模型，例如 CNN, RNN, KNN, 和 Logistic Regression。
3. **回测与评估**：  
   * 对模型进行回测，使用历史数据来验证每个模型的表现。

**表格结构**：

* + **行（Rows）**：每一行表示不同的 alpha 和 beta 组合。例如 (alpha=0.01, beta=0.02)，(alpha=0.02, beta=0.03) 等。
  + **列（Columns）**：每一列表示一个模型的表现。例如：CNN, RNN, Logistic Regression, KNN 等。
  + **单元格内容**：每个单元格中可以记录模型在该 alpha 和 beta 组合下的表现指标。你可以选择不同的表现指标，例如总收益（P&L 总和）、胜率、最大回撤（Max Drawdown）等。

| **alpha / beta** | **CNN P&L** | **RNN P&L** | **Logistic Regression P&L** | **KNN P&L** |
| --- | --- | --- | --- | --- |
| 0.01 / 0.02 | 1500 | 1600 | 1400 | 1300 |
| 0.02 / 0.03 | 1800 | 1700 | 1550 | 1450 |
| 0.03 / 0.04 | 1750 | 1650 | 1500 | 1400 |

* + **参数扫描（Grid Search）**：
    - 你可以采用类似于\*\*网格搜索（Grid Search）\*\*的方式，枚举一系列 alpha 和 beta 的组合，并在这些组合下对每个模型进行回测。
  + **模型比较**：
    - 对于每个 alpha 和 beta 组合，记录每个模型的总 P&L（或其他衡量标准）。
    - 观察在哪些参数组合下，模型的表现都比较好。这有助于找到相对稳定的参数组合，这些组合在不同的市场条件下都能表现出色。
    - 例如，如果在某些参数组合下，CNN 和 RNN 都有较好的 P&L，那么这可能是较为稳健的组合。
  + **选取最优组合，集成方法权重调整**：  
    - 在找到表现较好的参数组合后，可以进一步使用集成方法将不同模型的预测结果结合起来，最大化交易信号的准确性和收益。
    - 可以根据每个模型在表格中的表现（例如 P&L 总和）为其分配权重，表现较好的模型权重较大，以期望提高最终策略的收益稳定性。

买入信号构建方式

### **1. "基于止盈或止损条件" 逻辑：take\_profit or stop\_loss**

****data['buy\_signal'] = data.apply(lambda row: 1 if row['take\_profit'] or row['stop\_loss'] else 0, axis=1)

* **逻辑**：买入信号基于**止盈或止损**条件，意味着只要价格达到预期的上行目标（止盈），或者价格下跌到某个止损点（并可能存在反弹机会），就会产生买入信号。
* **影响分析**
  + **当日交易策略**：由于要求在**当天收盘前平仓**，你需要在当天就做出买入和卖出的决策。使用“or”逻辑意味着你只要看当天的价格达到设定的止盈或止损条件，就触发买入。
  + **高波动中的及时行动**：这种逻辑在有较大波动的市场中是有效的，可以在当天找到合适的买入机会并及时离场，符合不持有隔夜仓位的要求。
* **优缺点**：
  + **优点**：灵活的条件可以让你根据当天的价格波动迅速做出反应，这符合要求的“当天内完成交易”的条件。
  + **缺点**：如果市场当天的波动幅度不大，这种逻辑可能会导致没有买入信号，从而错失一些潜在机会。

### **2. 基于 P&L 的正负生成买入信号**

* **逻辑**：根据 P&L 的正负值决定是否买入，**P&L 为正**则产生买入信号，反之不买入。
* **影响分析**：
  + **历史表现局限性**：由于交易必须在**当天完成**，你无法根据历史累积的 P&L 来决定是否在当天进行买入。这种逻辑会显得比较滞后，因为你需要先有交易并计算其 P&L 才能判断是否进行下一次买入。
  + **不适合日内决策**：这种策略主要依赖于历史盈利的表现，因此在日内交易中使用会有滞后性，无法灵活地应对当日内的价格波动。
* **优缺点：**
  + **优点：可以减少盲目交易的次数，因为你只在有明确历史盈利的时候进行交易。**
  + **缺点：无法有效应对当天的市场波动，导致不适合当天开仓并平仓的要求，尤其是在波动性较大的市场中。**

### **3. "只要市场没有显现出明显的下跌趋势" 的逻辑（if not stop\_loss）**

****data['buy\_signal'] = data.apply(lambda row: 1 if not row['stop\_loss'] else 0, axis=1)

* **逻辑**：只要市场**没有触发止损**（即没有明显下跌），就认为有机会进入市场并生成买入信号。
* **影响分析**：
  + **过于乐观**：这种逻辑意味着在市场**没有明显下跌**时就买入，但这种策略没有考虑到当天的整体价格波动和潜在的短期回调，可能会让你在高点买入，然后不得不在同一天的低点平仓以防止损失。
  + **高频交易**：这种逻辑在没有显著下跌的情况下，会较为频繁地产生买入信号，这种过于积极的做法在日内交易中容易导致频繁交易且增加交易成本和风险。
* **优缺点**：
  + **优点**：可以抓住市场的小幅上涨机会，适合希望保持高频操作的投资者。
  + **缺点**：由于没有很好地考虑到当天内的波动和止盈点，可能会在市场横盘震荡时导致频繁的交易行为，不符合稳定日内交易的逻辑。

### **如何选择最合适的买入信号？**

#### **1. 根据市场情况和风险偏好**

* **市场波动较大且不确定时**：可以选择第一个 "or" 逻辑。因为它可以在\*\*市场表现极端（比如大涨或大跌）\*\*时灵活买入，适合高波动市场，目标是通过波动获利。
* **倾向于根据历史数据判断并进行保守决策**：选择基于 P&L 的逻辑。这种策略更适合希望有**数据支持、减少盲目决策**的情况，依赖历史表现来进行交易。
* **市场整体趋势向好、想保持积极买入**：选择第三种 "if not stop\_loss" 逻辑。这适合在市场没有显著下跌风险时进行操作，适合**牛市或稳健上涨**的市场。

#### **2. 测试与验证**

最好的做法是将这三种逻辑都实现并进行**回测**，然后根据回测结果来决定哪种逻辑最符合预期的表现。你可以通过以下步骤来选择：

* **回测每种逻辑**，观察它们在历史数据中的表现，尤其是 P&L 的总和、最大回撤（最大损失）、交易频率等指标。
* **比较收益与风险**：不仅要看收益，还要衡量风险和最大回撤。如果一种策略在高收益的同时伴随着高风险，那么可能需要适当调整。

### **建议的策略**

根据“不承担隔夜风险”的要求，最推荐的买入信号逻辑是**“or”逻辑**：

* **适合日内交易的灵活性**：
  + **"or" 逻辑**能够在当天价格达到某个目标值（无论是止盈还是止损）时迅速作出反应，这对于日内交易非常重要。
  + 在价格达到止盈目标时，可以快速入场并在短期内平仓；在价格快速下跌到止损点时，可以抓住可能的反弹机会，确保在当天内完成买入和卖出。
* **及时应对波动**：
  + GDX 的历史价格表现表明存在显著的波动性，而“or”逻辑可以帮助你在波动市场中寻找机会。
  + 由于策略要求在当天收盘前平仓，因此需要具备**足够的灵活性**来捕捉当天的市场变化，而不等待更长时间的确认信号。
* **控制交易频率**：
  + 使用“or”逻辑的一个好处是，你可以调整参数来控制交易的频率。
  + 例如，**通过调整止盈和止损的参数 (beta 和 alpha)**，你可以使交易条件更加严格或宽松，进而减少不必要的交易次数，优化你的日内交易决策。

### **进一步建议**

* **参数优化**：在设定“or”逻辑时，可以通过调整 alpha 和 beta 的值来优化买入信号。
  + 如果你发现市场波动很大，可以将 beta 设置得相对较高，这样只在价格涨幅较大时才买入，减少交易频率。
  + 同样，如果你想减少在市场剧烈波动下的交易次数，可以适当增大 alpha，避免频繁在下跌过程中买入。
* **回测验证**：建议对所有买入信号的逻辑进行回测，特别是验证在历史上不同市场状态下的表现。这样可以帮助你找到最适合日内交易且能带来最大化收益的逻辑和参数组合。

test set验证

根据结果，daily profit & loss 构建方式在多个模型中表现最好，可以考虑 **统一构建方式为 daily profit & loss**。至于 α,β，可以根据以下方法：

1. **全局优化 α,β\alpha, \betaα,β**：
   * 对所有模型的 Test Set P&L 取平均，寻找最优的 α,β组合。
2. **个性化模型设置**：
   * 如果某些模型在特殊 α,β 下表现突出，可以让它们独立优化，而其他模型保持统一。