**由于对于同一期彩票中的14场比赛，每场比赛地位等同，理论上互换后也不会影响预测结果，因此针对这个特点使用以下的模型架构**

**Shared Block and Final Block Architecture**

* **Purpose of Shared and Final Blocks**: The SalePredictor model is designed with a two-part architecture: a **Shared Block** and a **Final Block**, which together aim to extract shared features across multiple matches and then integrate these features for effective prediction. This architectural separation has several advantages in terms of parameter efficiency, feature representation, and predictive accuracy.
  + **Shared Block**:
    - The shared block is implemented using nn.Sequential and is applied **identically to each match**. It takes in **14 features per match** (such as odds and other match-specific data) and learns a shared representation across all 14 matches.
    - This block is used to process each match independently but with **shared parameters**. As a result, the model can learn common patterns or relationships that exist across different matches, such as how similar odds profiles might influence outcomes in similar ways regardless of the teams involved.
    - By reusing the same block across all matches, the model significantly reduces the number of trainable parameters, making it **computationally efficient**. This is particularly important when dealing with scenarios involving multiple repeated structures, like the 14 matches in the lottery.
  + **Final Block**:
    - After the shared block processes each match, the output features from all 14 matches are **concatenated** together along with other base features (such as statistical information over the past 30 issues).
    - The final block then takes these concatenated features and passes them through several layers to make the ultimate prediction. This block acts as a **specialized predictor** that is capable of utilizing both the match-specific extracted features and overall historical trends.
    - The final block is designed with multiple layers and outputs to predict the target variables (sale\_amount\_R9, prize\_count\_R9, sale\_amount\_14, prize\_count\_14\_1, prize\_count\_14\_2). These predictions rely on the interaction between the shared match-level features and the overall statistical characteristics of the lottery sales.
* **Benefits of This Architecture**:
  + **Parameter Efficiency and Generalization**:
    - The **shared block** approach greatly reduces the overall parameter count by using the same network to process all matches, as opposed to having separate blocks for each match. This helps prevent overfitting, especially when the data size is limited.
    - It also promotes **generalization**, as the network is encouraged to extract features that are broadly useful across different matches. This way, the model doesn't just memorize the specifics of each match but instead focuses on learning general patterns that could apply to any match.
  + **Modularity**:
    - The **two-block modular architecture** ensures that different parts of the model are specialized for different tasks. The shared block focuses on learning general match features, while the final block learns how to aggregate these features for meaningful predictions.
    - This modularity also makes the model easier to understand and maintain. It’s clear how each part of the model contributes to the prediction process—shared blocks focus on representation, while the final block focuses on integration and output.
* **Training Considerations**:
  + **Layer-wise Learning Rate**: The shared and final blocks are updated using separate optimizers with distinct learning rates. Specifically, the **shared block** uses a lower learning rate (0.0005), while the **final block** uses a higher learning rate (0.001). The lower learning rate for the shared block ensures that the features extracted remain stable and generalizable, while the higher rate for the final block allows it to quickly adapt to diverse match feature combinations for effective prediction.
* **Feature Combination and Learning Flow**:
  + During the forward pass, the shared block takes the match features, processes them, and produces an output representation for each match. These outputs are then **concatenated** to form a unified input to the final block.
  + The **final block** integrates these shared outputs, along with other historical statistical features, to predict the lottery sales and prize counts. The goal is to allow the model to focus on both **localized features** (specific to individual matches) and **global trends** (represented by overall statistics).

In essence, the combination of the **Shared Block** and **Final Block** enables the model to balance between learning **generic, reusable features** for each match and making **specific predictions** by combining these features intelligently. This separation into shared and specialized components makes the model more robust and helps it generalize better to new data.