B: Character RNN

In this part, I designed my own recurrent models in a new class called MyRNN.
The parameters I give to this class include the number of input_size, hidden_size, n_classes, num_layers and model. The explanation for these parameters are following:

- Input_size: the number of features in the input character.
- Hidden_size: the number of features in the hidden state h.
- Num_layers: number of recurrent layers, which means stacking how many RNN cells together.
- N_classes: the size of output.
- Model: whether we choose RNN or LSTM or GRU model

I trained for 200 epoches, but the performance won't increase and I cannot find the problem.

C: Linear Quadratic Regulators

1. $x_{t+1} = A x_t + B u_t$

cost function: $x_t^T Q x_t + u_t^T R u_t$

cost to go: $J_{t+1}(X) = x^T P_{t+1} X$

$J_t(x) = \min_u (x^T Q x + u^T R u + J_{t+1}(A X + B u))$

To find minimum u, set the gradient w.r.t u equal to 0,

$2 u^T R + 2 (A X + B u)^T P_{t+1} B = 0$.

$u^* = -(R + B^T P_{t+1} B)^{-1} B^T P_{t+1} \cdot A X$

$J_t(X) = x^T Q x + u^{*T} \cdot R u^* + (A X + B u^*)^T P_{t+1} (A X + B u^*)$

$= x^T (Q + A^T P_{t+1} A - A^T P_{t+1} B (R + B^T P_{t+1} \cdot B)^{-1} B^T P_{t+1} A / 2$

$P_t = Q + A^T P_{t+1} A - A^T P_{t+1} B (R + B^T P_{t+1} B)^{-1} B^T P_{t+1} \cdot A$

or $P_t = Q + K_t^T R K_t + (A + B K_t)^T P_{t+1} (A + B K_t)$

$K_t = -(R + B^T P_{t+1} B)^{-1} B^T P_{t+1} \cdot A$.