

关于GRU的维度：首先隐层会和r、z做Hadamard乘积，所以必定以隐层的维度为目标，故而在GRU内部，Tensor均转化为Hidden Size，理论上有6个矩阵，工程实现上只有两个矩阵，因为有3个矩阵与Xt想乘，3个矩阵与隐向量相乘。

对于GRU而言，input size namely the size of Xt don't have to equal to hidden size,cause matrix Wwill change the input size to hidden size

out\_put also have the same dimmention with hidden size

in short hidden size is everything

python 文件命名以字母开头，如果以数字开头的话无法import

对于单层GRU而言：

# 输入：seq\_len，batch\_size，input\_size

# 隐层：layers, batch\_size，hidden\_size

# 输出：seq\_len，batch\_size，hidden\_size

# 隐层：layers, batch\_size，hidden\_size