

Faça como eu fiz

Na etapa de classificação de nossa prática, utilizamos uma GRU com uma única camada e uma camada linear para classificar a feature temporal produzida pela GRU. Dessa forma, ao alimentar a camada linear, utilizamos o seguinte comando:

```
output = self.linear( hidden.squeeze() )
```

[COPIAR CÓDIGO](#)

A função `squeeze()` aplicada à memória interna irá eliminar todas as dimensões iguais a 1. Dessa forma, a dimensão original da memória, $(L \times B \times F)$ sendo $L = 1$ o número de camadas, $B=64$ o tamanho do batch e $F=256$ o tamanho da feature, será transformada em $(B \times F)$.

Faça como eu fiz na aula, dessa vez definindo uma GRU com duas camadas. Como ficará a passagem da memória interna da GRU para a camada linear?