



正向和反向的RNN的W不同，所以初始化隐藏层的数量要翻倍
并且两个层的输出是concat在一起的（连起来）

PYTHON

d21 中初始化隐藏层部分

```
def begin_state(self, device, batch_size=1):
    if not isinstance(self.rnn, nn.LSTM):
        # nn.GRU以张量作为隐状态
        return torch.zeros((self.num_directions * self.rnn.num_layers,
                             batch_size, self.num_hiddens),
                             device=device)
    else:
        # nn.LSTM以元组作为隐状态
        return (torch.zeros((
            self.num_directions * self.rnn.num_layers,
            batch_size, self.num_hiddens), device=device),
                torch.zeros((
                    self.num_directions * self.rnn.num_layers,
                    batch_size, self.num_hiddens), device=device))
```