Activation Functions

1. Usually, what activation function would you use if you had more than 10 hidden layers? **1 / 1 point**

   ○ Sigmoid

   ○ Tanh

   ⦿ Relu

   > ✓ **Correct**
   > correct

2. What is the problem with the tanh and sigmoid activation function? **1 / 1 point**

   ○ They are discontinuous functions

   ○ You can't take the derivative

   ⦿ The derivative is near zero in many regions

   ○ They are periodic functions

   > ✓ **Correct**
   > correct

3. What activation function is used in the following class **1 / 1 point**

```python
class NetRelu(nn.Module):
    def __init__(self,D_in,H,D_out):
        self.linear1=nn.Linear(D_in,H)
        self.linear2=nn.Linear(H,D_out)
    def forward(self,x):
        x=torch.relu(self.linear1(x)))
        x=self.linear2(x)
        return x
```

   ⦿ relu

   ○ tanh

   ○ Sigmoid

   > ✓ **Correct**
   > correct