Short Answer

1. What is the best slope you could find?
Slope = 0.1 seems the best slope. The accuracy and loss after 20 epoch is 0.964 and 0.124 respectively. (Actually there is not so much difference between different slope.
What happens if you set the slope > 1?
If I set slope = 2, the accuracy and loss after 20 epoch is 0.961 and 0.127 respectively. It seems that it doesn't matter if slope>1.
What about slope < 0?
If I set slope = -1, the accuracy and loss after 20 epoch is 0.963 and 0.125 respectively. It seems that it doesn't matter if slope<0.
what happens if you set slope = 1?
If I set slope = 1, the accuracy and loss after 20 epoch is 0.922 and 0.276 respectively. Obviously the result is not good enough. If slope = 1, the activate function becomes y=x, so the input and output between each layer is linear, as a result, all the layer can combined as one linear layer.

2. Set PReLU to take 1 slope per layer. After 20 epochs, what were your PReLU slopes? Does this correspond with what you found in question 1?
I have two PreLU layer and the first prelu went 0.5386while the second prelu went -0.8936. The final accuracy is 0.964 and loss is 0.112. I think it matches what I found in question1. The slope can be negative or >1, but it can't equal to 1

3. If you add more layers and more epochs, what accuracy can you reach? Can you get to 99%? What is your best network layout?
I add 5 linear layers and after 50 epoch, the accuracy became 0.9751, the loss is 0.112. Unfortunately I didn't get to 99%, maybe I need more flexible net structure or other layers, instead of just using linear layer.