1. What is the role of topological sort in the backpropagation algorithm?

**Topological sort will have the effect of getting all the nodes in an order that will can use the backward function with respect to the leaf nodes.**

2. What is a pitfall that can happen when computing the gradients of a variable that depends on multiple other variables (i.e., what mistake did Andrej make in this case?)

**If you have multiple variable gradients, you need to add together all the other gradients together to cumulate the gradients so we actually get the correct gradient at the end without messing up any of previous work**

3. What mistake did Andrej make when he was propagating gradients?

**He was storing the return from the functions in self.\_backward but not the actual functions.**