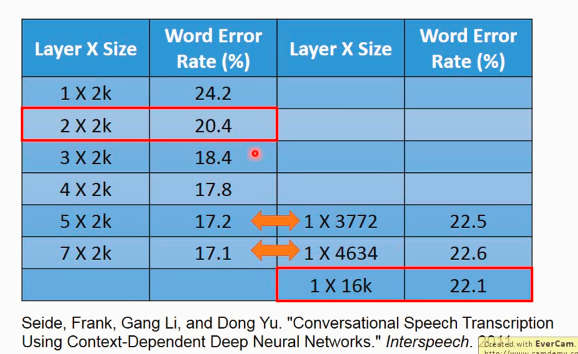
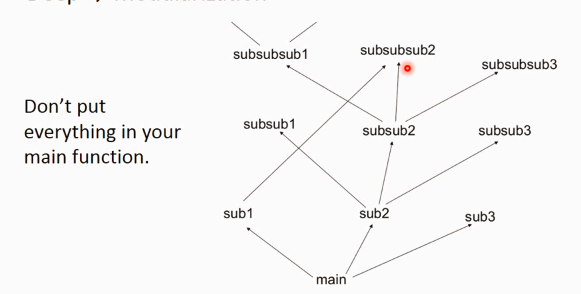
Why deep? (why not wide)

Deep learning is not the method/case that when you have more parameters, it gets better result. You can see from the table below that if you make it wider not deeper, even the parameters of wide network is more, the result is worse.

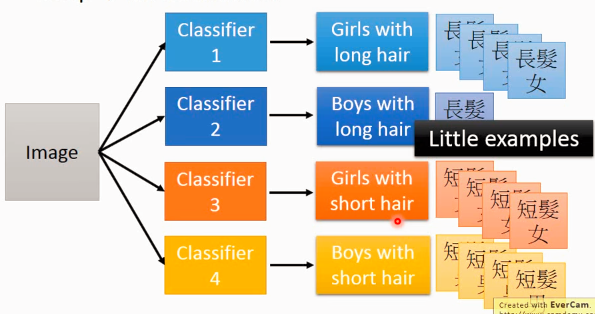


Deep learning is more like modularization. Think about programming:



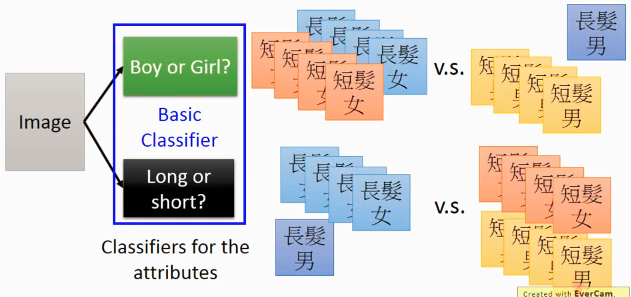
If you organize your function like this(C:\Users\SHIYAO~1\AppData\Local\Temp\SGPicFaceTpBq\13088\013D4A80.png), some functions are used in different functions(weird sentence ε=(´ο｀\*))),let’s see an example). E.g. subsubsub2 is sorting function, it’s used by sub1, subsub2 and so on, we just need to call it, instead of write it every time we need it.

So, in machine learning, supposing we want to do this classification

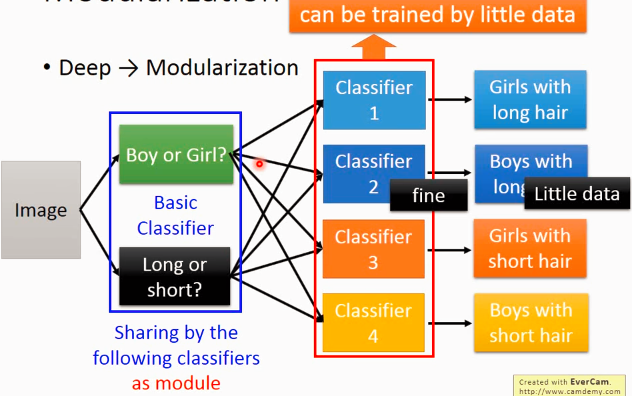


Since there are few examples in class “boys with long hair”, so Classifier 2 might be weak.

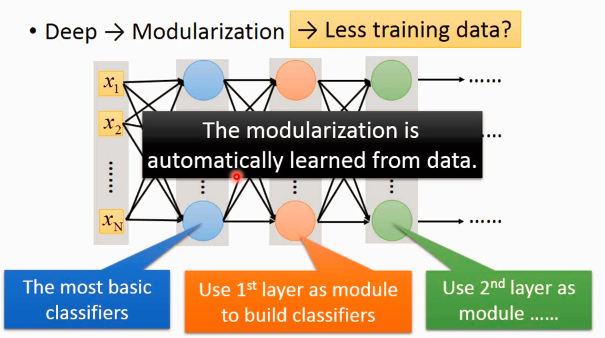
So we modularize like this to make each basic classifier can have sufficient training examples:



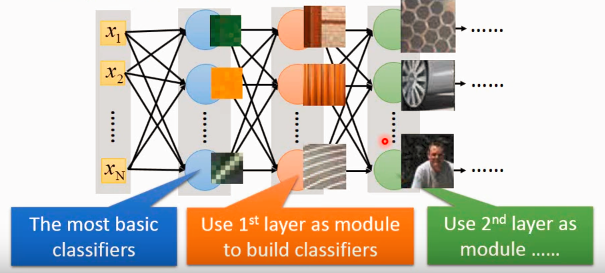
Then, for the final classification use:



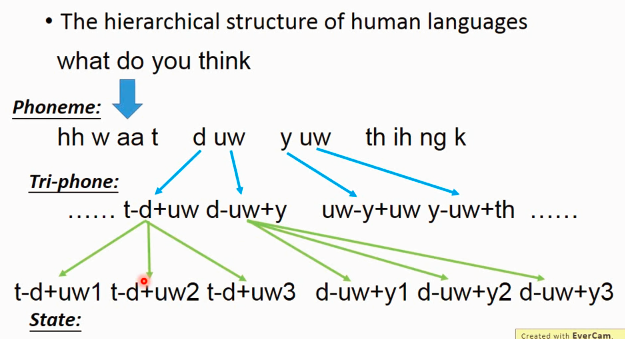
How this related to deep learning?



One example on image:



Modularization works good auch in Speech.



Tri-phone does not mean the combination of 3 phones. When the same phone in different context, we use different models to represent it.

