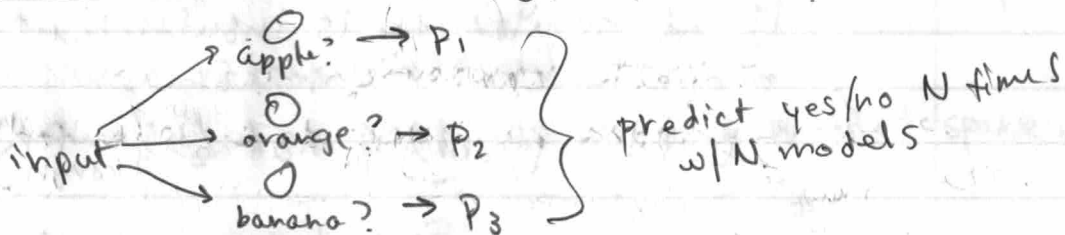


# MULTI-CLASS NEURAL NETS

Logistic regression works well for binary classification.  
What about multi-class problems?

E.g. is this an apple, banana, orange

One-vs-all:  $N$  log reg nodes, 1 per class



Can also use a neural net w/ layers that share internals - pretty efficient to train.

Softmax: Add a constraint: sum of probabilities per-class must sum to 1 (encode this in the loss fn.)

Which one?   
- single label } Softmax  
- multiple labels } one vs all,  $N$  log reg (logit)

Softmax options

- full: brute force, calc. for all classes (expensive)
- candidate sampling: calculate for positive labels, but only random sample of negatives (i.e. at training time)