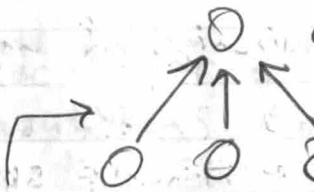
NEURAL NETWORKS

Ideally our model would learn the nonlinearity instead of manual feature cross definition

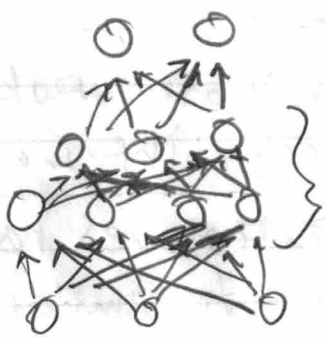
Linear model



Adding additional linear combination layers doesn't add nonlinearity!

Activation function: Some simple but nontrivial nonlinear function — just adds nonlinearity to an input/output relation!

E.g. tanh
sigmoid
RELU (rectified linear unit) = max(0, x)



linear combinations w/ learned weights

nonlinear activation fn. on result

Omg it just learns weights for well-behaved noise generators

Neural net =
- set of nodes (neurons) in layers
- set of weights to connect layers
- set of biases, one per node
- activation fn. to transform layer outputs (usually by layer)

Tuning

- # of layers adds redundancy to handle nondeterminism
- # of nodes in a layer increases the range of nonlinearity that can be modeled
- diff. activation fns. may be more efficient on diff. problems
- too many layers/nodes can overfit or memorize the training data