Parameters vs Hyperparameters

Parameters: $W^{[1]}, b^{[1]}, W^{[2]}, b^{[2]}, W^{[3]}, b^{[3]} \dots$
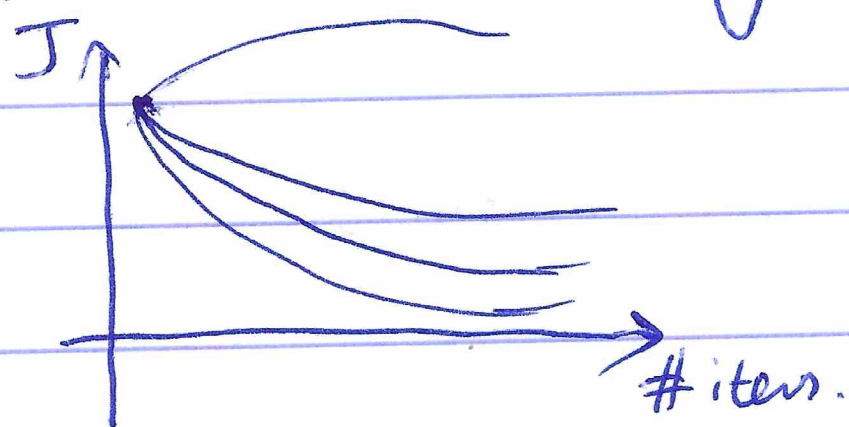Configuration variables that are internal to the model and whose value can be estimated from the data

Hyperparameters: α, #iterations, ~~minibatch~~ L, $n^{[1]}, n^{[2]}, \dots$, activation no.?
Configurations that are external to the model and whose value cannot be estimated from the data
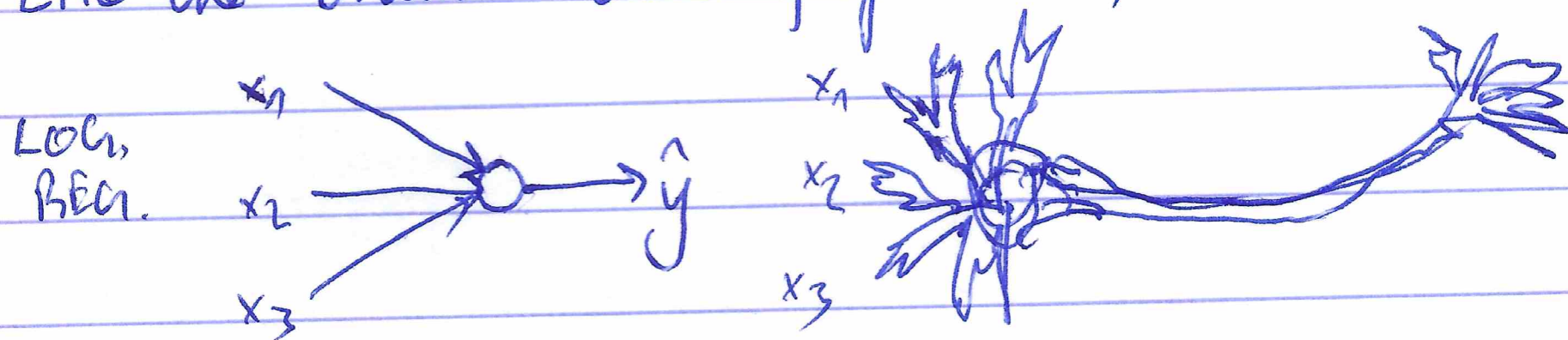
Later: momentum, minibatch size, regul., ...

- Applied DL is a very empirical process

Relation of DL to Brain

- "Like the brain": oversimplification, but there is a simple analogy

LOG.
REG.

Week 4 Quiz Notes

- "Cache": We use it to pass variables computed during for. prop. to the corresponding backprop. step. It contains useful values for backprop. to compute derivatives

- The deeper layers are typically computing more complex features of the than the earlier layers