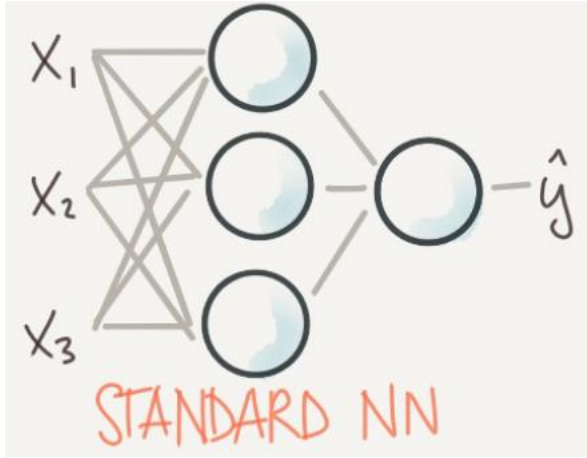
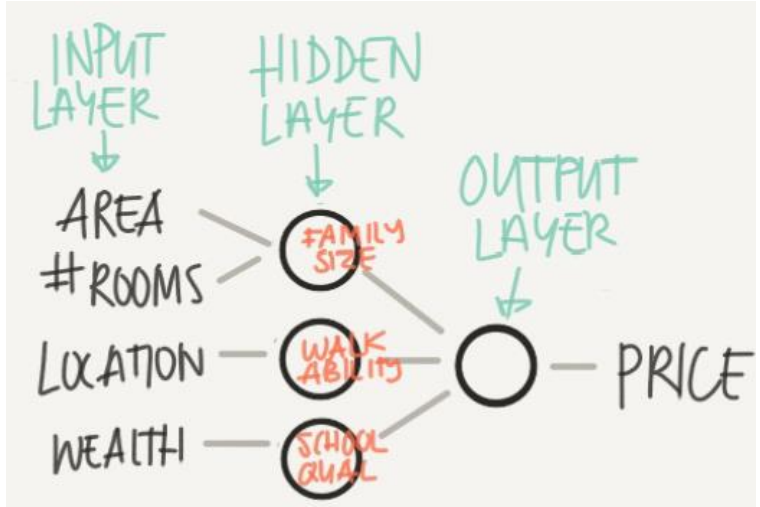
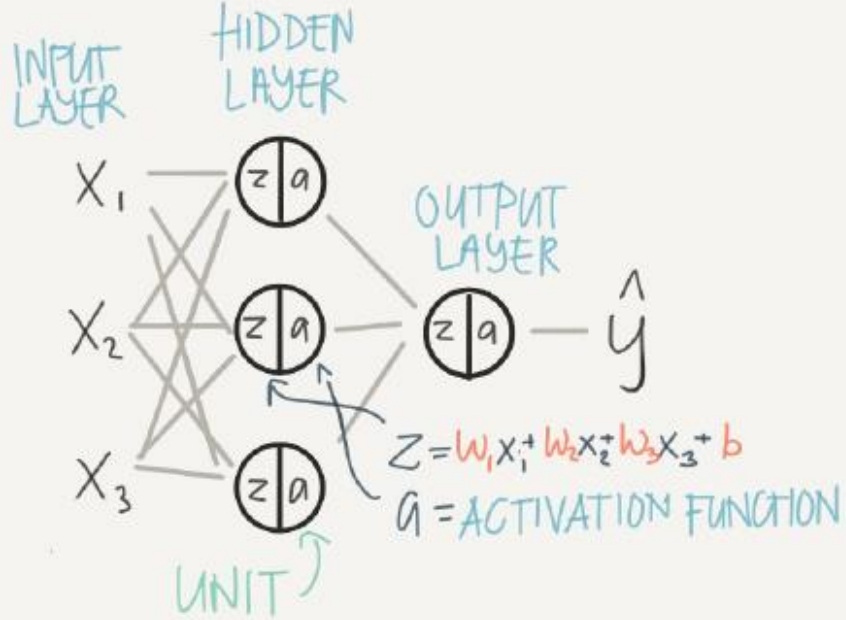
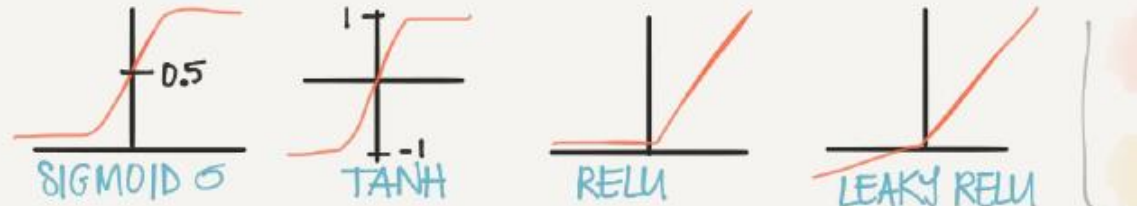




2 LAYER NEURAL NET



ACTIVATION FUNCTIONS



BINARY CLASSIFIER
- ONLY USED FOR OUTPUT LAYER

SLOW GRAD DESCENT SINCE SLOPE IS SMALL FOR LARGE/SMALL VAL

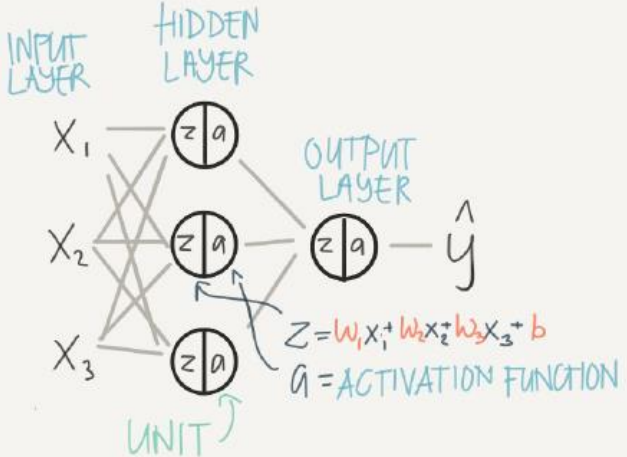
NORMALIZED ⇒ GRADIENT DESCENT IS FASTER

DEFAULT CHOICE FOR ACTIVATION

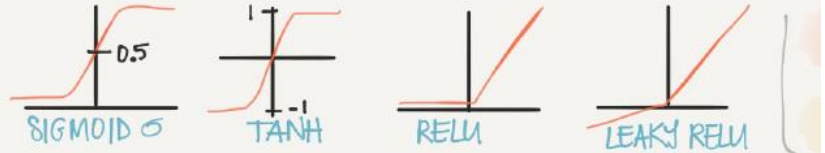
SLOPE = 1/0

AVOIDS UNDEF SLOPE AT Ø BUT RARELY USED IN PRACTICE

2 LAYER NEURAL NET



ACTIVATION FUNCTIONS



BINARY CLASSIFIER - ONLY USED FOR OUTPUT LAYER

SLOW GRAD DESCENT SINCE SLOPE IS SMALL FOR LARGE/SMALL VAL

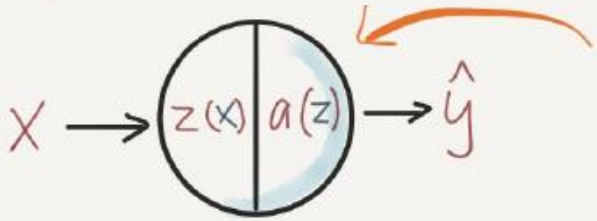
NORMALIZED ⇒ GRADIENT DESCENT IS FASTER

DEFAULT CHOICE FOR ACTIVATION

SLOPE = 1/0

AVOIDS UNDEF SLOPE AT Ø BUT RARELY USED IN PRACTICE

THE TASK IS TO LEARN w & b



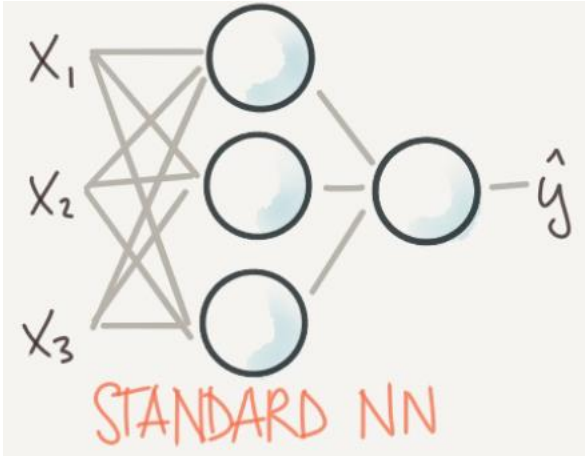
MINI NEURAL NET

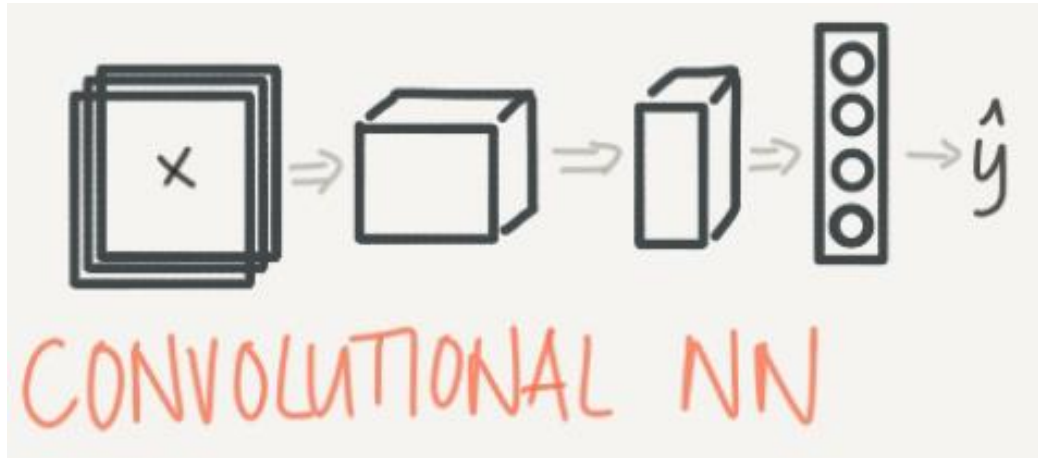
$$z(x) = wx + b$$

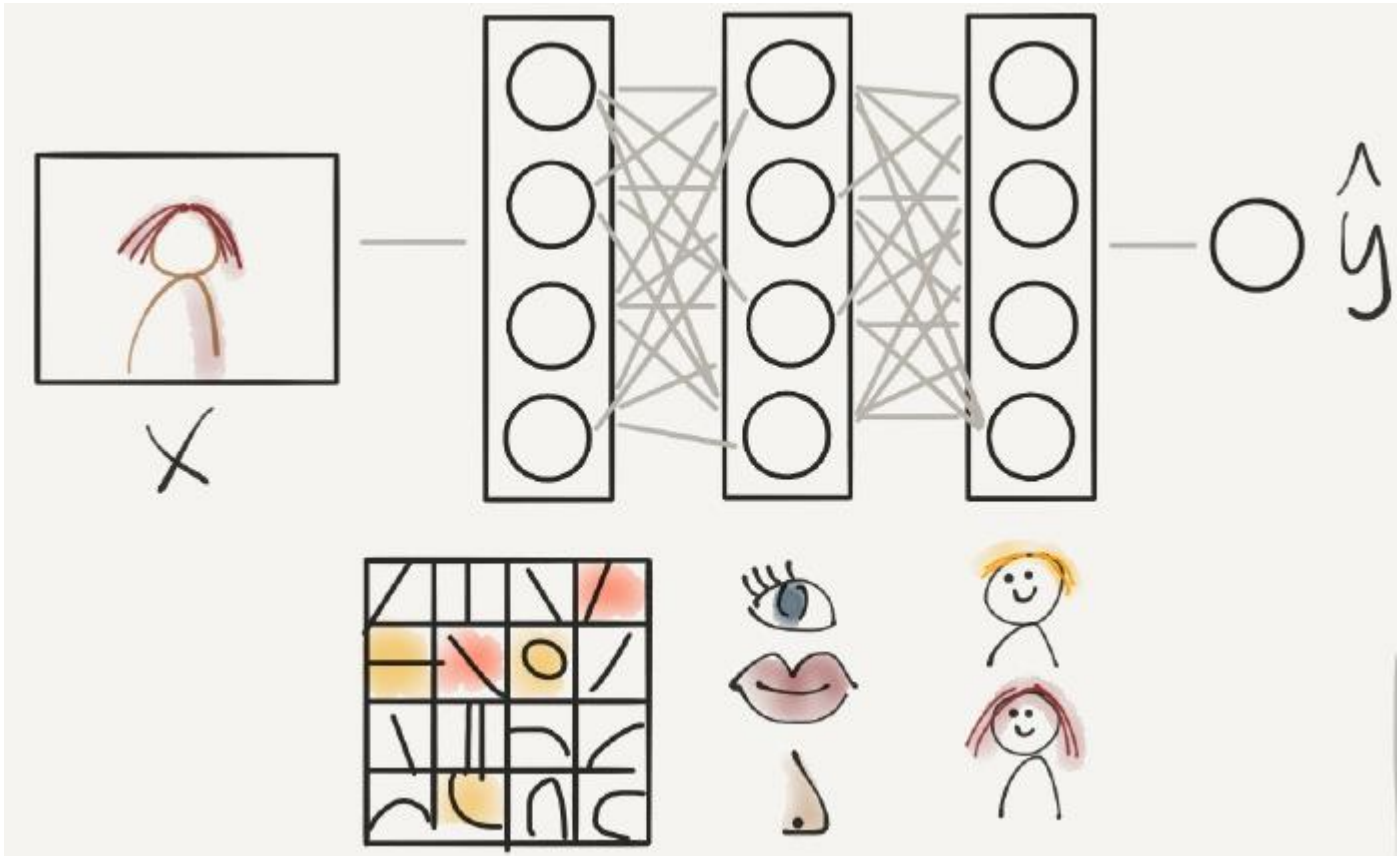
$$\hat{y} = a(z) = \sigma\,\text{SIGMOID}(z)$$

1. FORWARD PROPAGATION • CALCULATE $\hat{y}$
2. BACKWARD PROPAGATION • GRADIENT DESCENT + UPDATE w & b

REPEAT UNTIL IT CONVERGES


x_1
x_2
x_3
$\hat{y}$
STANDARD NN


x
$\hat{y}$
CONVOLUTIONAL NN

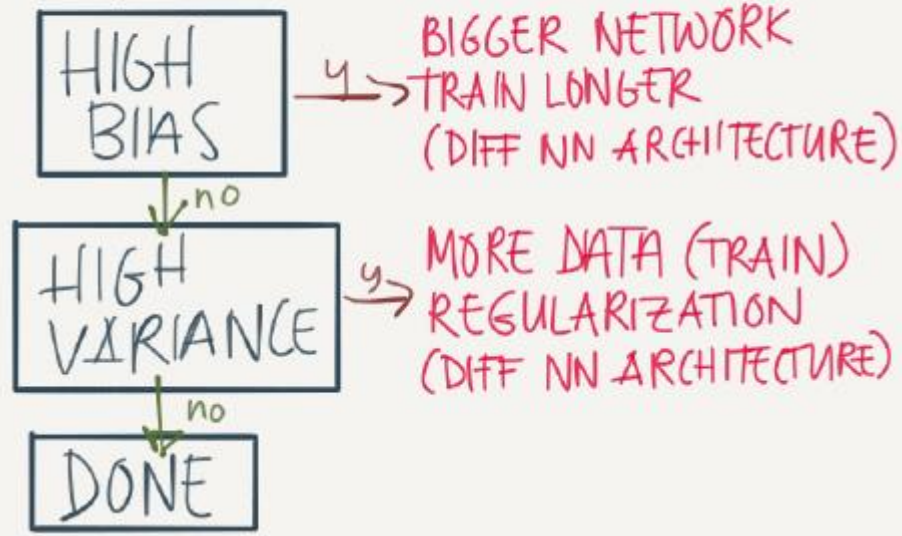

x
$\hat{y}$

BIAS/VARIANCE

	ERROR			
TRAIN	1%	15%	15%	0.5%
TEST	11%	16%	30%	1%
	HIGH VARIANCE	HIGH BIAS	HIGH BIAS & VARIANCE	LOW BIAS & VARIANCE

ASSUMING HUMANS GET 0% ERROR

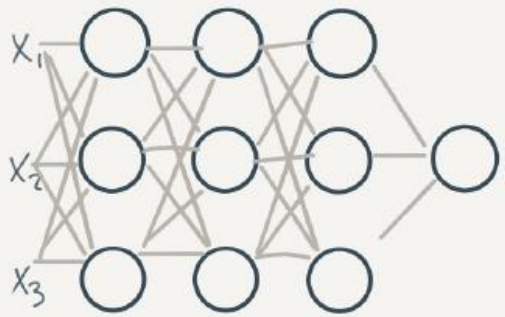
THE ML RECIPE

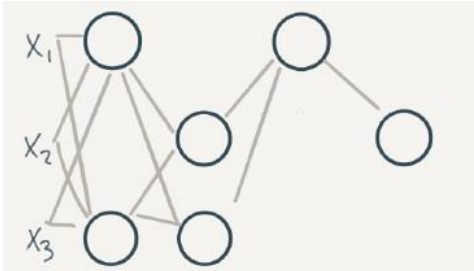


REGULARIZATION

PREVENTING OVERFITTING

DROPOUT

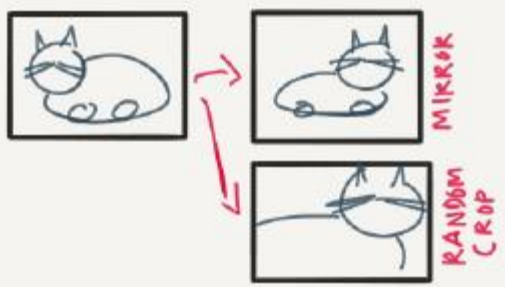




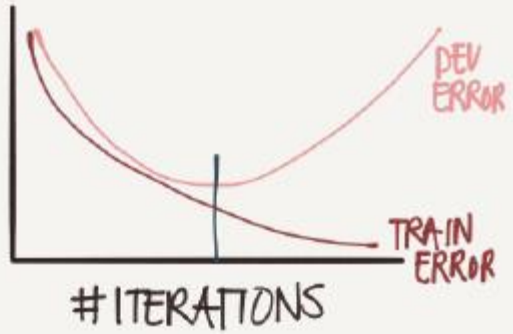
WE GET SIMPLER NWs & LESS CHANCE TO RELY ON SINGLE FEATURES

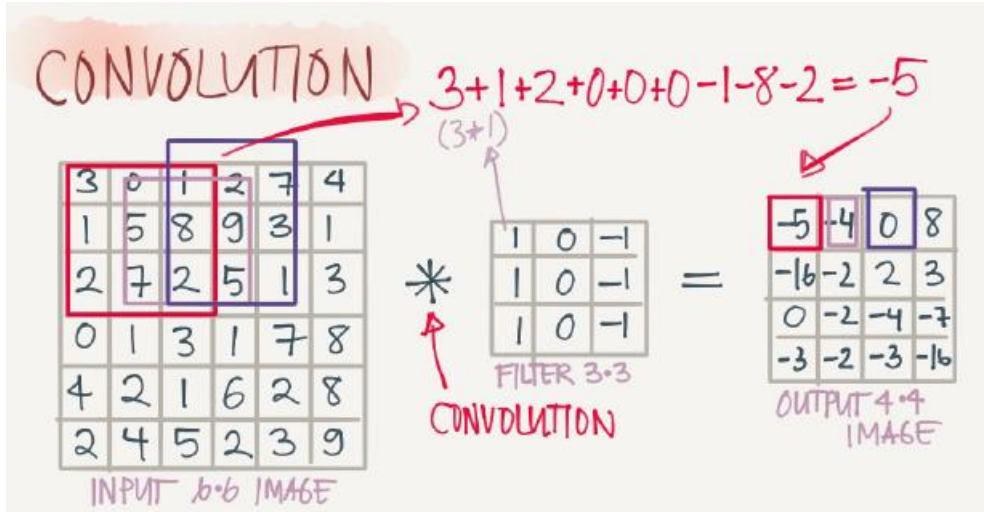
DATA AUGMENTATION

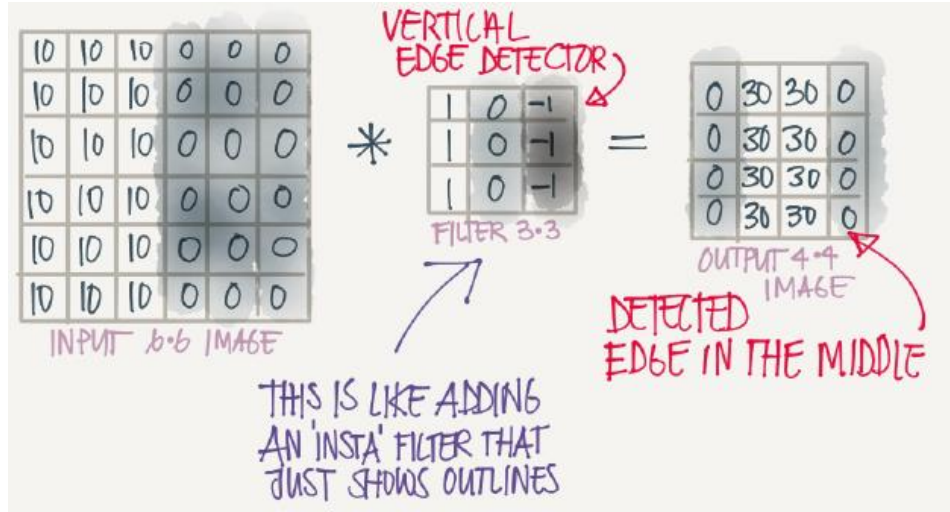
GENERATE NEW PICS FROM EXISTING



EARLY STOPPING







WE COULD HARD-CODE FILTERS. JUST LIKE WE CAN HARD-CODE HEURISTIC RULES.... BUT.... A MUCH BETTER WAY IS TO TREAT THE FILTER # AS PARAMS TO BE LEARNED

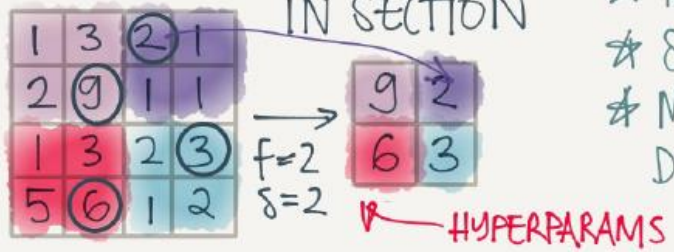
w_1	w_2	w_3
w_4	w_5	w_6
w_7	w_8	w_9

TYPICAL CONV.NET LAYERS

CONVOLUTION
POOLING
FULLY CONNECTED

POOLING (MAX)

FIND MAX VAL IN SECTION

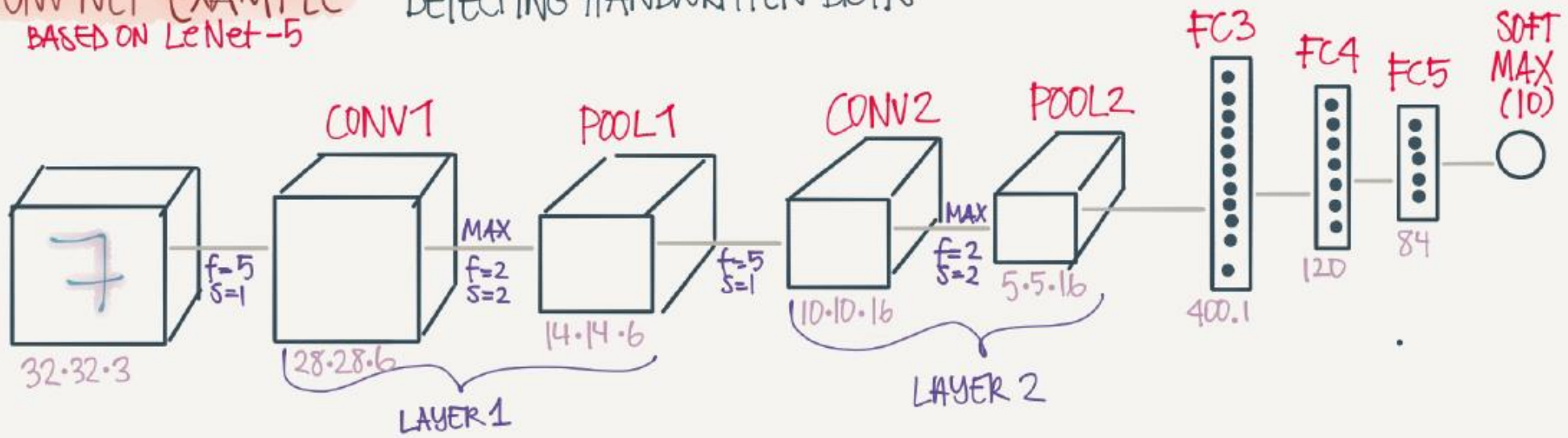


- REDUCES SIZE OF REPRES.
- SPEEDS UP COMPUTATION
- MAKES SOME OF THE DETECTED FEAT. MORE ROBUST

CONV NET EXAMPLE

BASED ON LeNet-5

DETECTING HANDWRITTEN DIGITS



DATA AUGMENTATION

WE ALMOST ALWAYS NEED MORE DATA TO TRAIN ON

USE OPEN SOURCE IMPLEMENTATIONS

SOME OF THE PAPERS ARE HARD TO IMPLEMENT FROM SCRATCH - USING OS YOU CAN REUSE OTHER PPLS WORK

DON'T FORGET TO CONTRIBUTE

TIPS FOR DOING WELL ON BENCHMARKS/COMPETITIONS

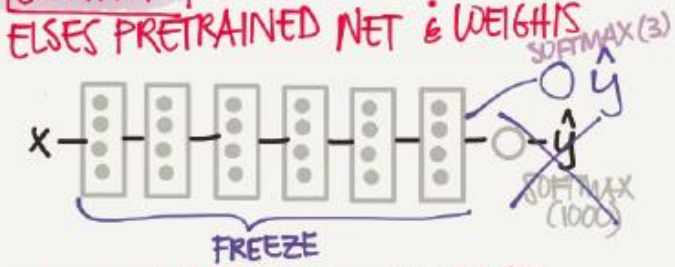
* ENSEMBLING.
  AVG OUTPUTS FROM MULT NN
* MULTI-CROP AT TEST TIME
  AVG OUTPUTS FROM MULTIPLE CROPS OF THE IMAGE

IN PRACTICE THEY ARE NOT USED IN PRODUCTION BECAUSE THEY ARE COMPUTE & MEM EXPENSIVE

TRANSFER LEARNING

WANT TO TRAIN A CLASSIFIER FOR YOUR CATS BUT DON'T HAVE ENOUGH PICTURES

SOLUTION DOWNLOAD SOMEONE ELSES PRETRAINED NET & WEIGHTS



FREEZE THE PARAMS, AND JUST REPLACE THE SOFTMAX LAYER WITH YOUR OWN & TRAIN

IF YOU HAVE MORE PICS - RETRAIN A MORE OF THE LATER LAYERS (MAYBE INITIALIZING WITH THE PRETRAINED WEIGHTS)