

TRAINING

Uses backpropagation

- Forward: run net, update each layer's input weights to get desired output from prev. layer using error derivative from...
- Back propagation of error derivatives from the output of the net back through its layers (uses chain rule etc. to combine derivatives)
- Uses dynamic programming to save computations of intermediate results on forward/backward passes

Tensor Flow handles this internally

Note: Requires differentiable functions!
Avoid too much depth!
Don't go too fast!

} Vanishing gradients
} Exploding gradients
} Dead ReLU units

Normalization helps

- reasonable scale, e.g. (-1, 1)
- linear scaling, log, caps
- avoid outliers

Dropout regularization: with prob. p , drop out nodes at random.