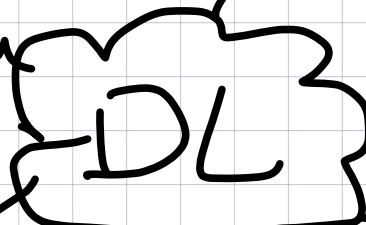


Common Practices

- Test Data
- Learning Rate
- Hyper Params?
- Ensembling
- Class Imbalance
- Evaluation



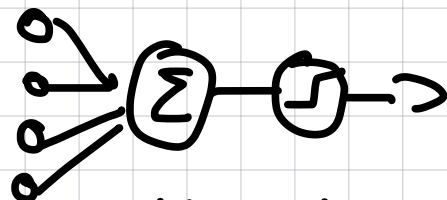
Introduction

- Deep Networks
- Amazing Result
- Factors:
 - GPU
 - Activations
 - Open Source
 - Open Data

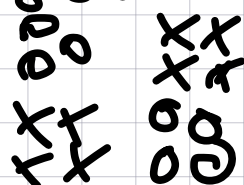
Neural Networks

- Rosenblatt Perceptr.

$$\hat{y} = \text{Sign}(w^T x)$$



=> Can't solve XOR



- Universal Approx. Theorem

$$F(x) = \sum_i v_i \phi(w_i^T \vec{x})$$

$$|F(x) - f(x)| < \epsilon$$

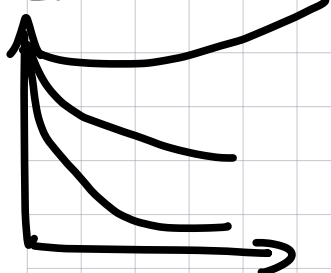


X

-> Tree Example

Activations + (NN)

- Linear
- Sigmoid
- Tanh
- ReLU
- LReLU
- ELU / SELU?
- Gradient + Loss Curves



Validation Loss?

CNN:

- Spatial Context
- Shared Weights
- $E_{e-1} = W^T E_e$
- $\nabla W = F_c X^T$
- Pooling?

Training: $L(\vec{x}, \vec{y})$

- Backprop
- Chain rule
- Debugging

$$L(W_1, W_2, W_3) = \frac{1}{2} \|W_3 W_2 W_1 \vec{x} - \vec{y}\|_2^2$$

$$\frac{\partial L}{\partial W_2} = (W_3 W_2 W_1 \vec{x} - \vec{y})(W_2 W_1 \vec{x})^T$$

$$\frac{\partial L}{\partial W_2} = W_3^T (W_3 W_2 W_1 \vec{x} - \vec{y})(W_2 W_1 \vec{x})^T$$

$$\frac{\partial L}{\partial W_1} = W_2^T W_3^T (W_3 W_2 W_1 \vec{x} - \vec{y}) \vec{x}^T$$

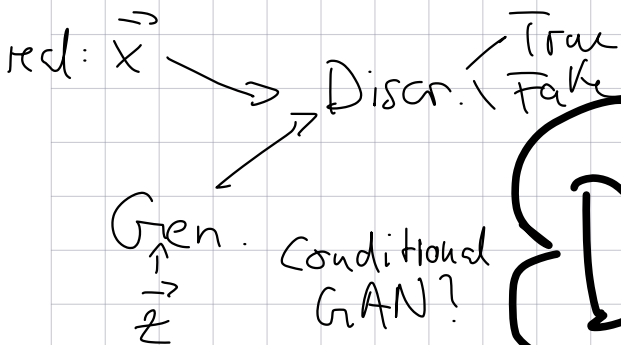
Loss + Optimization

- Univariate Gaussian -> L_2 Loss
- Multinoulli -> Cross Entropy Loss
- Hinge Loss + SVM
- SGD + Mini Batch
- Momentum + Nesterov
- ADAM?

Unsupervised

- Autoencoders?
- Denoising AE?
- Stacked AE?
- Sparse AE?
- GAN

New: Variational AE!

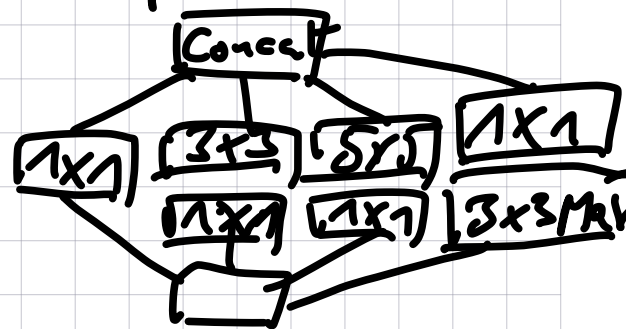


Architectures

- LeNet: Conv + Pool + Conv + Pool + FC + FC
- AlexNet + GPU + ReLU + Dropout + Mini Batch SGD
- Network in Network
- 1x1 Conv - Global spat. Pooling
- VGG:

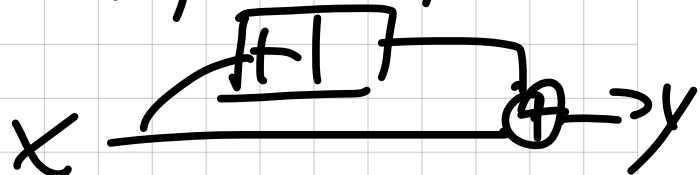


- Google-Net
- Inception Block



- ResNets

$$F(x) = f(x) + x$$



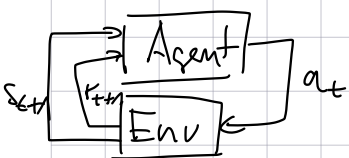
- DenseNet

Visualization

- Architecture
- Training
- Params
- Kernel
- Activations
- Occlusion
- Maximally Activating Patches
- t-SNE
- Backprop for Vis
- Optimizing Input

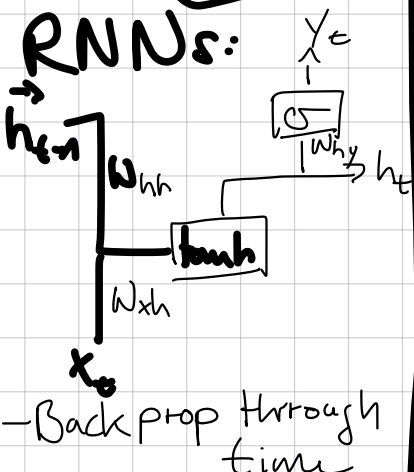
New: Attention!

- Reinforcement Learning
- Sequential Decision Making
- Markov Decision Process



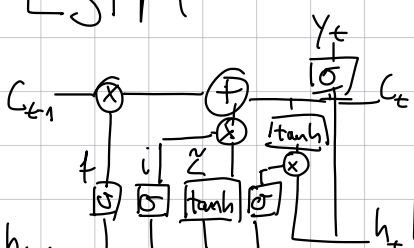
- Random Policy
- Greedy Policy
- State Value Function?
- Action Value Function?
- Deep Q Learning?

RNNs:

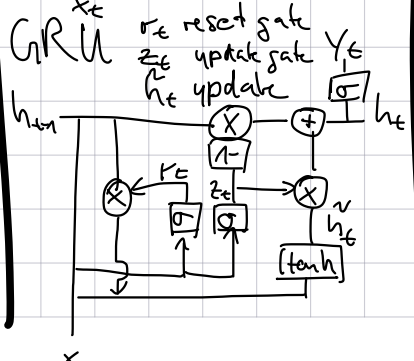


- Backprop through time

LSTM:



GRU:

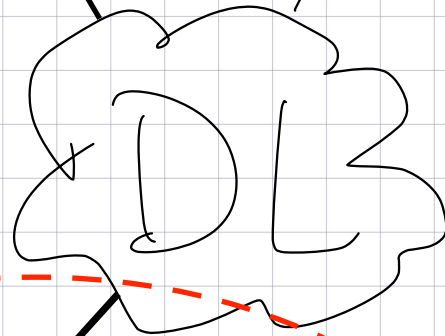


Graph Deep Learning

- The Graph Laplacian
- Spatial Aggregation

Weakly and Self-Supervised Learning

- From Class Label to Bounding Boxes
- From Bounding Box to Pixel Segmentation
- Self-Supervised Labels



Object Detection

- Region-based
 - Regional CNN
 - Fast R-CNN
 - Faster R-CNN
- Single Shot
 - YOLO
 - Retina-Net

Segmentation

- Fully Conv
 - Encoder
- Upsampling
 - Decoder
- Context Knowledge
 - Long Skip Connections
 - U-Net
- Adversarial Loss
- Instance Seg?

Prior Knowledge

- Known Operators (Precision Learning)
- Variational Networks?
- Deriving Deep Architectures
 - Video

Not Part of the Exam