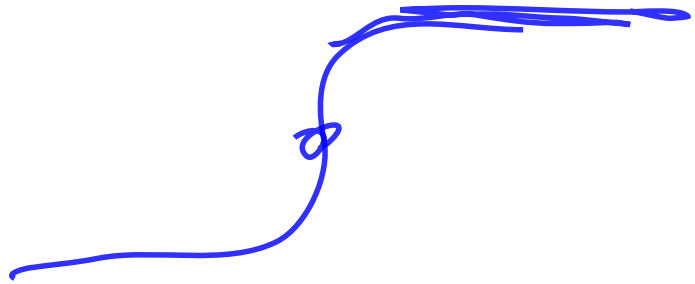
SMAI-M20-L33: Intro. to Feedback Networks; and Clustering

C. V. Jawahar

IIIT Hyderabad

November 6, 2020

Class Review



1. What is the a good initialization? (or what is a bad initialization[1])
2. What happens during training for an MLP?

Xavier

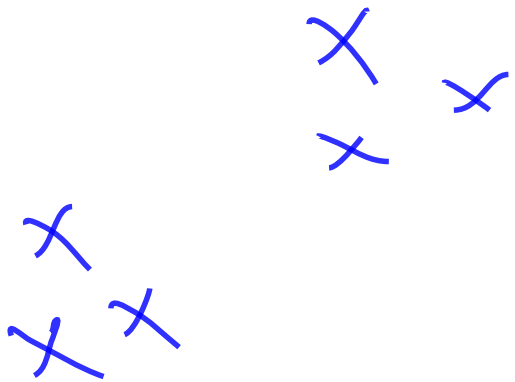
[1] Read, Tryout and Appreciate: https://www.deeplearning.ai/ai-notes/initialization/

Blank

Recap:

- **Matrix Factorization and Applications:** (i) SVD, (ii) Eigen Decomposition (iii) Matrix Completion (iv) LSI (v) Recommendations
- **Dimensionality Reduction and Applications:** (i) Feature Selection and Extraction (ii) PCA (iii) LDA (iv) Eigen face
- **Supervised Learning:** Formulation, Conceptual Issues, Concerns etc. (i) Loss Functions and Optimization (ii) Probabilistic View, Bayesian View, MLE (iii) Eigen Vector based optimization (iv) Gradient Descent: Stochastic and Batch GD (v) Classification and Regression
- **Classifiers:** (i) Nearest Neighbour, (ii) Notion of a Linear Classifier (iii) Perceptrons (iv) Bayesian Optimal Classifier (v) Logistic Regression (vi) Multiclass classification architectures (v) Decision Trees (vi) SVMs (hard margin, soft margin, kernel) (vii) Kernel trick and kernelized algorithms
- **Neural Network Architectures and Learning** (i) Neuron model, Single Layer Perceptrons (ii) SLP(iii) MLP (iv) Backpropagation (v) Chain rule (vi) Activations (vii) challenges in optimization (viii) Momentum (ix) Convolutional Layer (x) Auto-encoder and unsupervised learning
- **Beyond Simple Supervised Learning** (i) Paradigms of Learning

This Lecture:

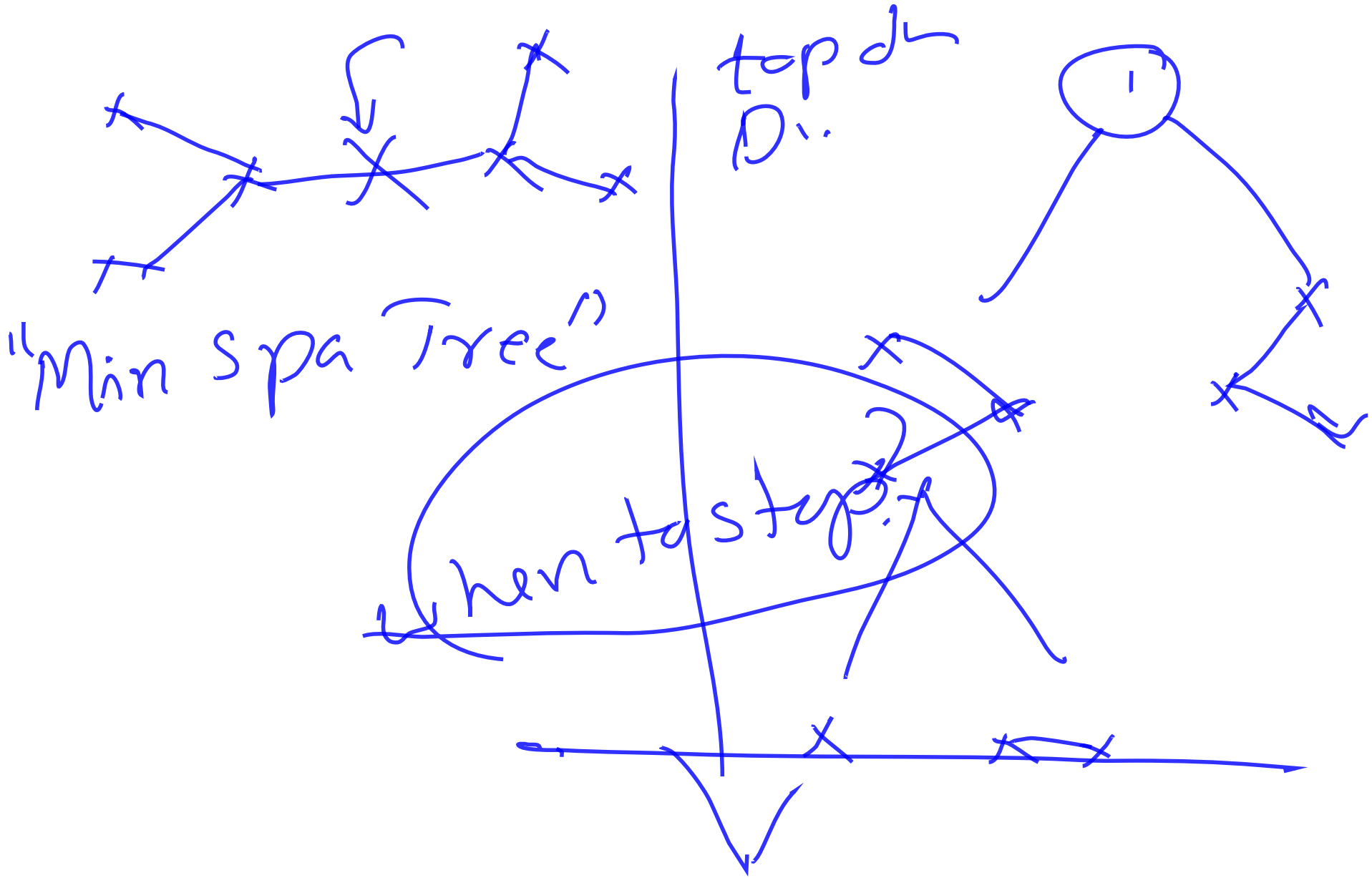


1. **Feedback/Recurrent Networks**
   1. Feedforward vs Feedback network
   2. Data as a sequence (of vectors)
   3. Recurrent Model
2. **Problem of Clustering**
   1. Notion of a Cluster (vs Class in supervised Learning)
   2. Hirarchical Approaches
   3. Agglomerative and Divisive (bottom-up and top-down)
3. Reading Material for Neural Networks [2]

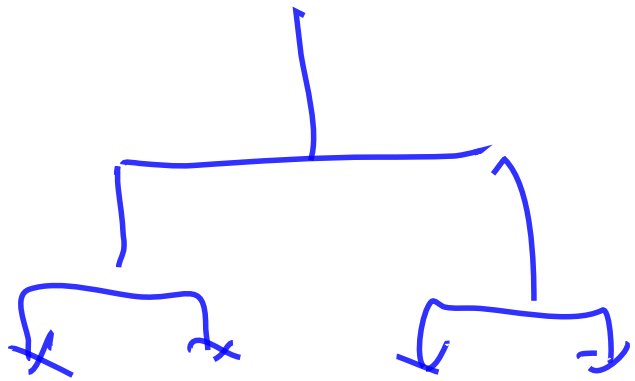
Questions? Comments?

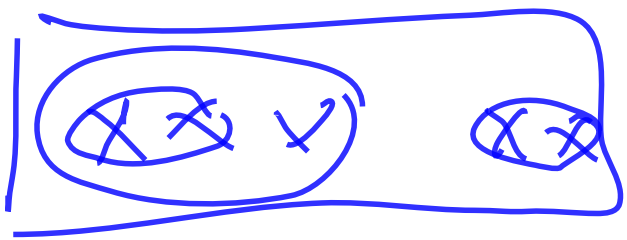
[2] Read in detail: https://www.dropbox.com/s/g9vu0dollo6sr48/nn-notes.pdf?dl=0

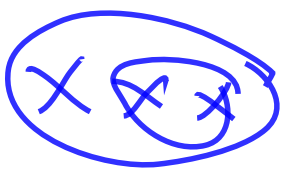
Blank



Blank







Blank

Discussions Point - I

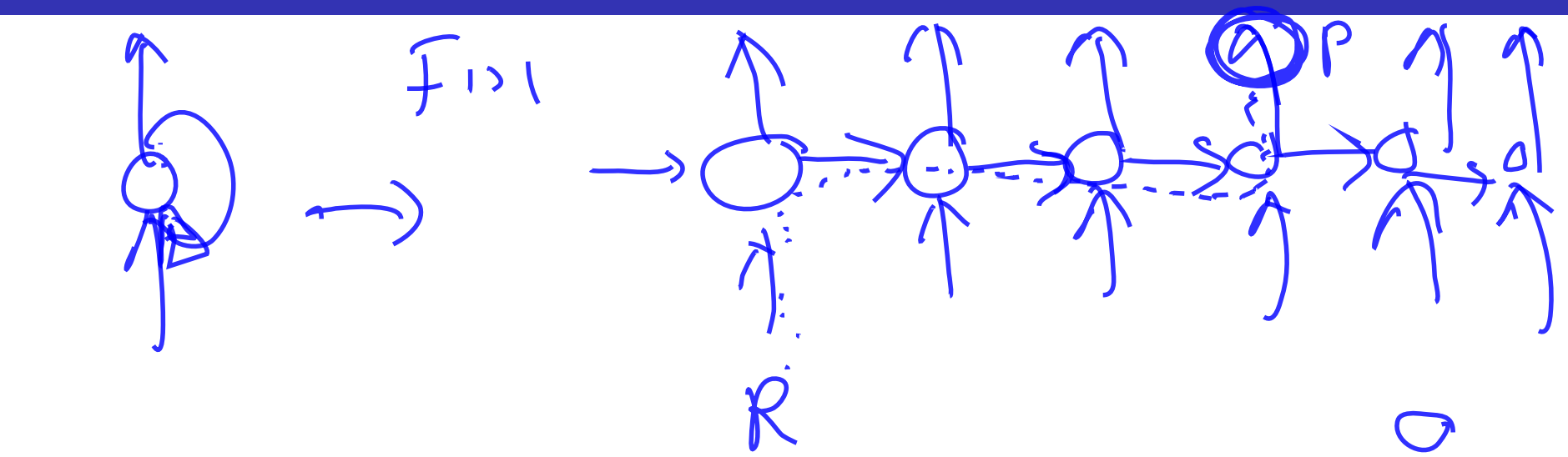
Consider a recurrent/feedback model of neural network given by

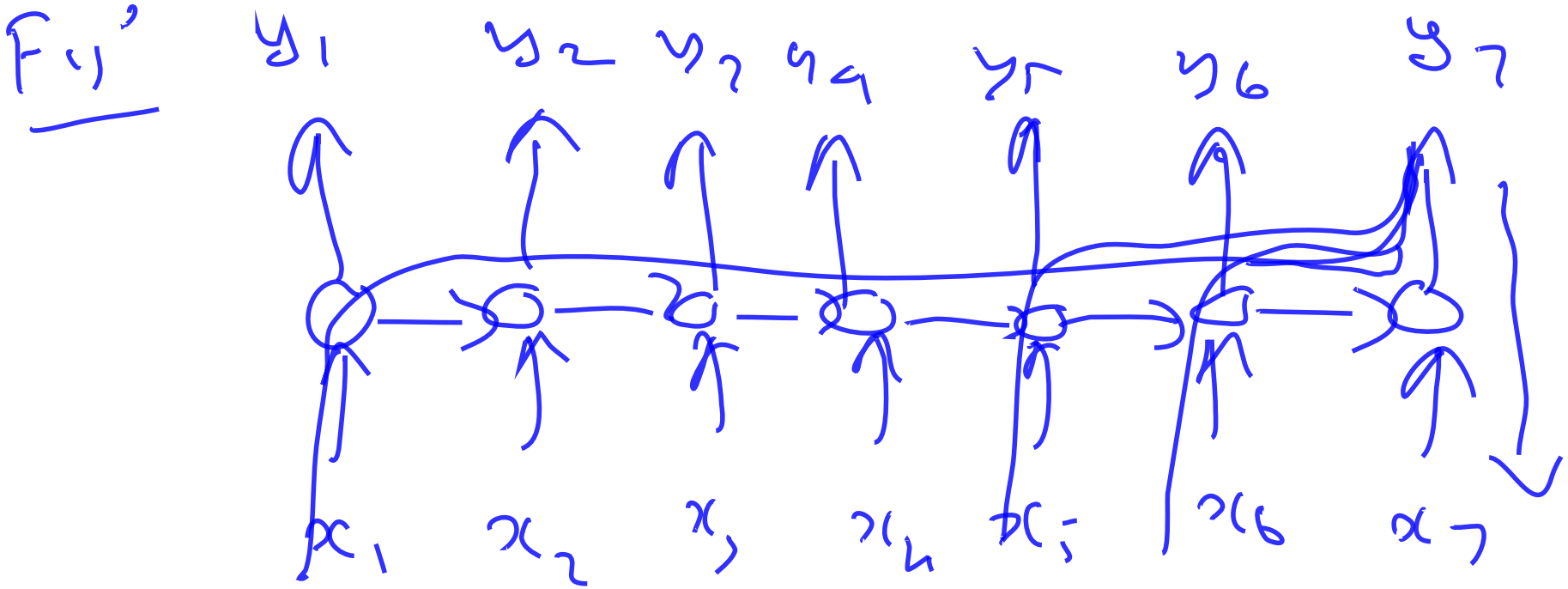
$$s_t = f(Ux_t + Ws_{t-1})$$

$$o_t = g(Vs_t)$$

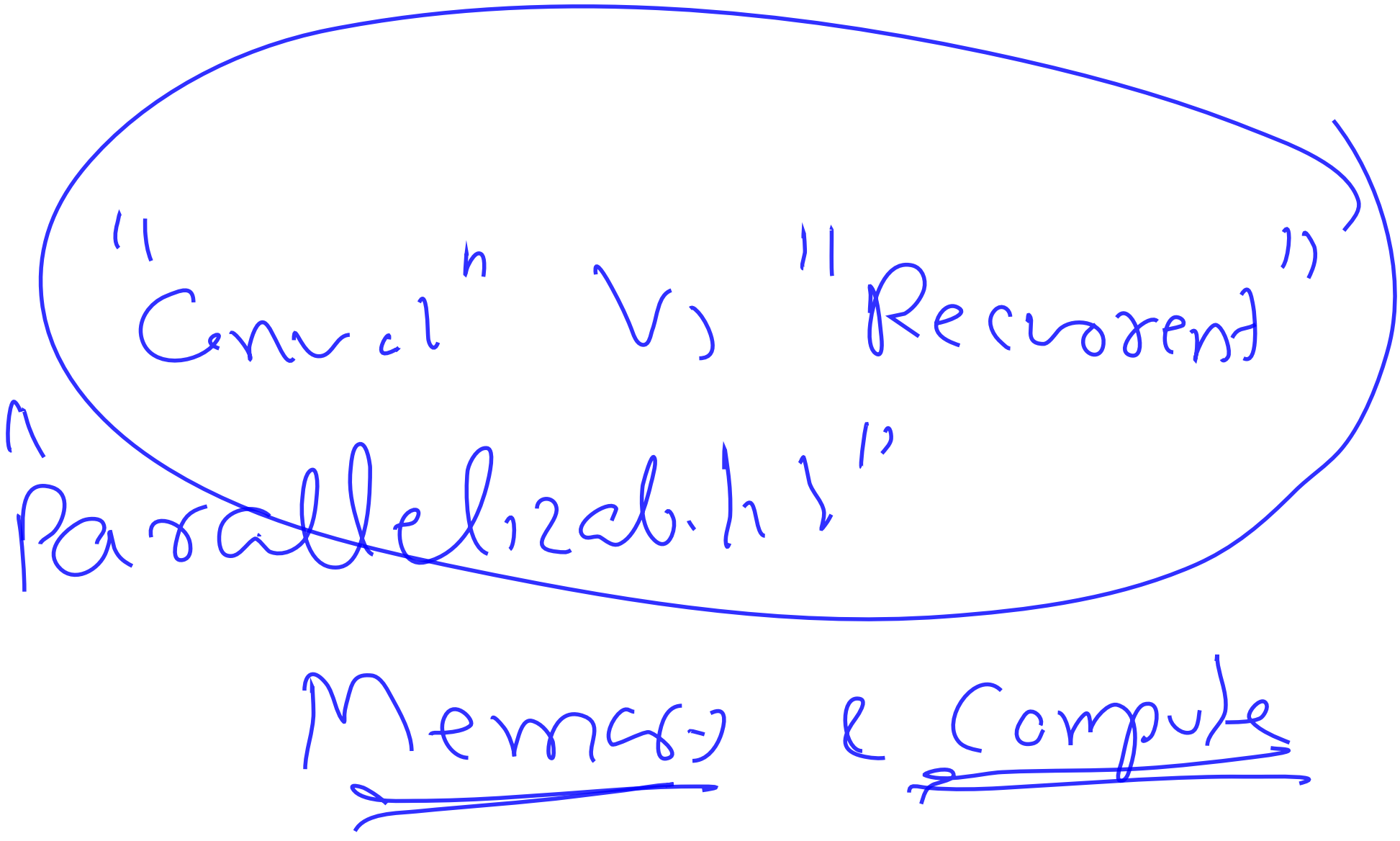
1. Why do one say that such networks have "infite or long term memory"? (or never forgets) Or RNNs capture long term dependencies?
2. Why the problem of vanishing gradient is serious in RNNs?

Blank





Blank



Blank

Discussions Point -II

Consider two Clusters $\mathcal{C}_1$ and $\mathcal{C}_2$. As part of merging/comparing clusters (say while designing a bottom-up or agglomerative clustering) we want to compare or find similarity between clusters. What do you think of the following functions[3]?

1. $\min_{x_1 \in \mathcal{C}_1, x_2 \in \mathcal{C}_2} s(x_1, x_2)$
2. $\max_{x_1 \in \mathcal{C}_1, x_2 \in \mathcal{C}_2} s(x_1, x_2)$
3. $Average_{x_1 \in \mathcal{C}_1, x_2 \in \mathcal{C}_2} s(x_1, x_2)$

Any alternate suggestions?

$C_1 \longleftrightarrow C_2$

mean & median

Medoid

[3] https://towardsdatascience.com/understanding-the-concept-of-hierarchical-clustering-technique-c6e8243758ec

Blank

Blank

Blank

Blank

What Next:?

1. NN Architectures and NN Learning (winding up)
2. Programming for Deep Learning.
3. Beyond Simple Supervised Learning