SMAI-M20-L32: Unsupervised Learning in Neural Networks

C. V. Jawahar

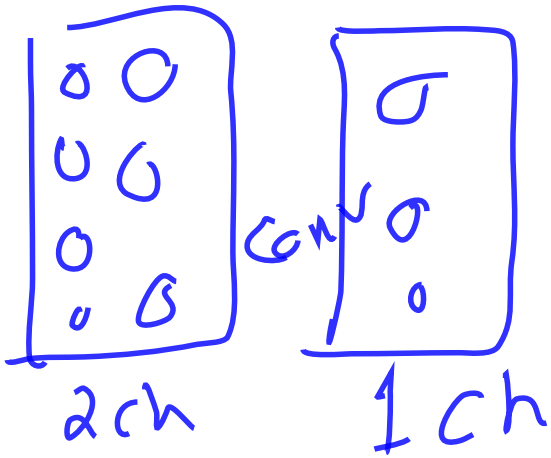
IIIT Hyderabad

November 2, 2020

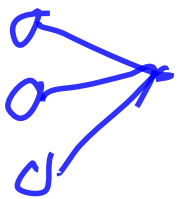
Announcement

1. Quiz in the regular class slot on Wed. Similar to Q1/CR.
2. Topics: Topics that are not covered in Q1.
3. NN Learning (included): MLP and Back Propagation
4. NN Learning (not included): Convolution layer, Auto encoder, Momentum

Class Review



1. Convolution layer in 1D CNNs:
   1. M channels in the input and N channels in the output.
   2. Number of learnable weights or parameters
   3. Stride and impact on output size
   4. Zero padding and impact on output size.



Blank

Recap:

- **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
- **Dimensionality Reduction and Applications:** (i) Feature Selection and Extraction (ii) PCA (iii) LDA (iv) Eigen face
- **Matrix Factorization and Applications:** (i) SVD, (ii) Eigen Decomposition (iii) Matrix Completion (iv) LSI (v) Recommendations
- **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

This Lecture:

1. **Auto Encoders**
   1. Reconstruct itself with a constrained ("bottleneck") architecture.
   2. Role as:
      1. Data Compression
      2. Unsupervised Feature Learning
      3. Non-Linear Dimensionality Reduction
   3. Comment about Popular Encoder-Decoder architectures of today.
2. **Beyond Supervised Learning**
   1. Supervised Vs Unsupervised Learning
   2. Unsupervised as "Clustering", "Discovery of the structure"
   3. Semi-Supervised Learning
   4. Self-Supervised Learning

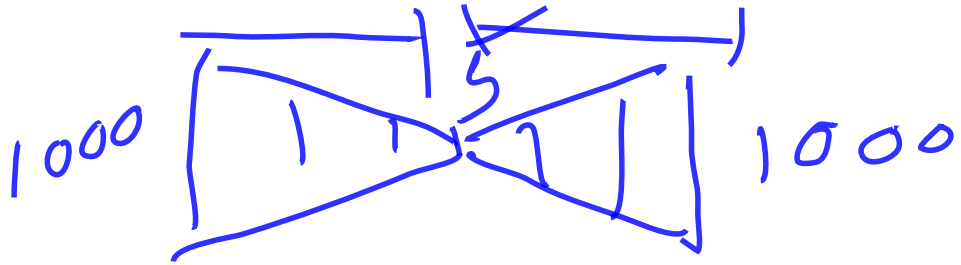
Questions? Comments?

Blank

Blank

Blank

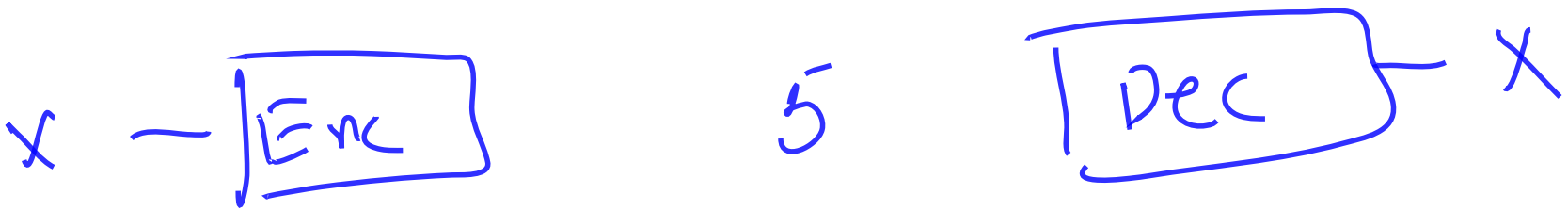
Discussions Point - I



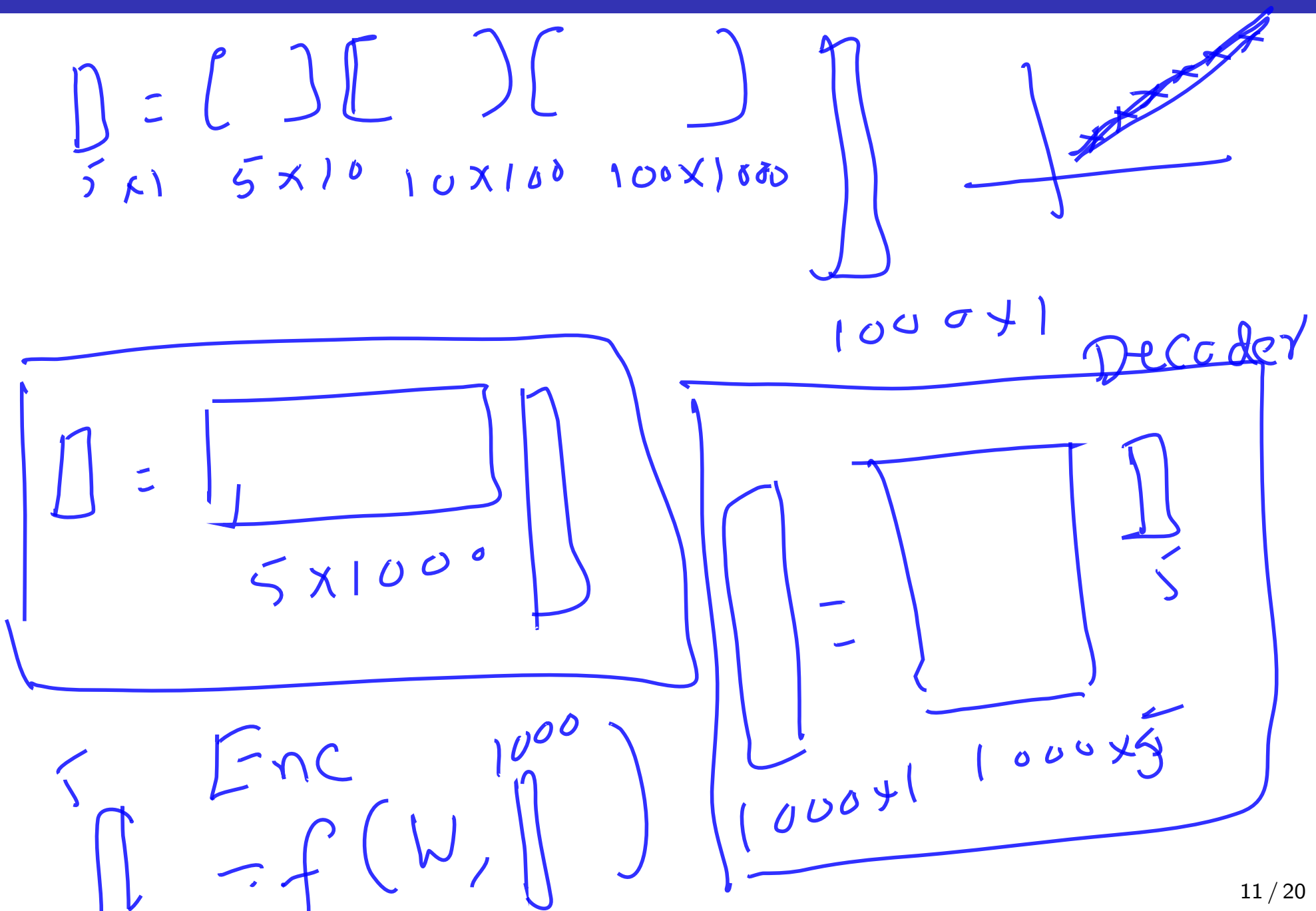
Consider an auto enoder with fully connected layer and the architecture as:

$$1000 - 100 - 10 - 5 - 10 - 100 - 1000$$

1. Assume the activations are linear, show how this is similar to "PCA" or linear dimensionality reduction that we are familiar with.
2. If we use it for compression (say for a speech or image signal of size 1000), what can we say about the compression ratio? Why is this a good idea? Why is this a bad idea?



Blank


5x1
5x10
10x100
100x1000
1000x1
Decoder
5x1000
1000x1
1000x5
Enc
1000
=f(W,

Blank

Blank

Discussions Point -II

We know the learning rule as:

$$w^{k+1} \leftarrow w^k - v_k$$

$$v_k = \eta \nabla J + \beta v_{k-1}$$

1. What if $\eta = 1.0$ and $\beta = 0.0$?
2. What if $\eta = 0.0$ and $\beta = 1.0$?
3. Assume $\nabla J = -0.1$ for all $k = 0, 1, \ldots, 10$. $v_{-1} = 0$. $w^0 = 0.0$. $\eta = 0.1$ and $\beta = 0.9$, what happens to w^k for $k = 1, 2, 3$?
4. Where it should have reached with and without momentum for $k = 10$? (Appreciate how momentum helps in speeding up, if we have a consistent slow slope).

Blank

Blank

Blank

Blank

What Next:?

1. NN Architectures and NN Learning
2. Programming for Deep Learning.
3. Beyond Simple Supervised Learning