22. Approximate Data Deletion via Weight Pruning

- **Report Link**: PDF
- **Year**: Spring 2021

1. Problem Statement & Formalization

Enable “right to be forgotten” by approximating removal of training data via model pruning.

- **Input**: trained model parameters
- **Output**: pruned model approximating deletion of specific data
- **Metrics**: retention accuracy vs baseline, privacy leakage

2. Data / Dataset

- **MNIST**, **CIFAR-10** datasets used in experiments.
- Links:http://yann.lecun.com/exdb/mnist/ , https://www.cs.toronto.edu/~kriz/cifar.html

3. Model Design / Methods

- Baseline: retraining from scratch without target data
- Advanced: weight pruning, knowledge distillation

4. Training & Validation

- Train CNN on MNIST/CIFAR-10
- Apply pruning simulating deletion
- Validate retained accuracy vs retraining

5. Evaluation & Reporting

- Accuracy before/after deletion
- Measure leakage with membership inference tests

6. Extensions / Bonus

- Explore structured pruning vs unstructured
- Extend to large models (ResNet)

7. Deliverables

- Pruning code
- Dataset scripts
- Report comparing pruning vs retrain baseline