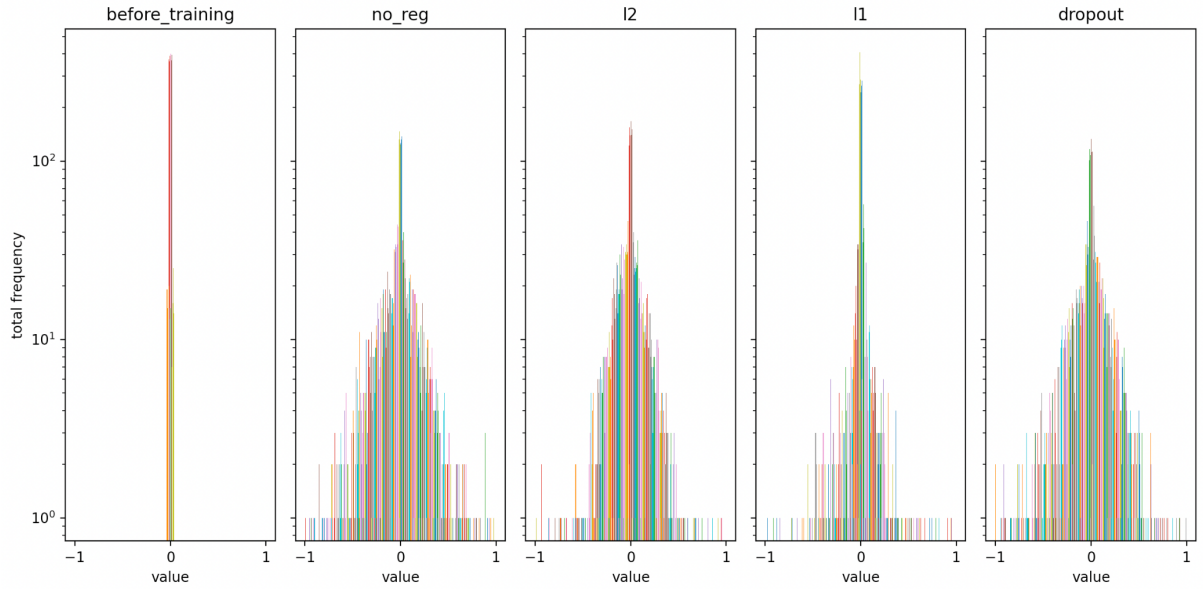
**Submission – Assignment 5**

Marta Gulida – matriculation number 5585808 – mg776
Erik Bode – matriculation number 4505199 – kb301
Tillman Heisner – matriculation number 4517815 – th273

## Pen and Paper Task

Parameter Distribution




L1 has more zeros than L2. Because if we visualize the regularizer L2 will have a shape of the circle, and L1 will have the shape of the square with corners aligned to x and y axes.
Without no regularization the magnitude of the weights are not punished as such they are larger weights than the regularized ones.
L2 punishes the magnitude of the weights than L1 as such there are more smaller weights.
For dropout the model effectively becomes an ensemble of many sub-networks, each with different sets of dropped-out connections. It introduces noise or randomness.

**Data Augmentation**. Data augmentation increases model generalization by augmenting the training set with fake data.

1) *State five effective operations to generate fake data on the MNIST dataset.*
    - Rotation. Rotate the digit images by small angles (up to 15 degrees).
    - Scaling. Slightly scale the images, making them larger or smaller.
    - Brightness and Contrast Adjustment.
    - Horizontal and Vertical Shifts. Shift the images horizontally and vertically by small amounts.

- Gaussian Noise. Add random Gaussian noise to the pixel values of the digit images.

2) *State one operation which doesn't make sense and why.*

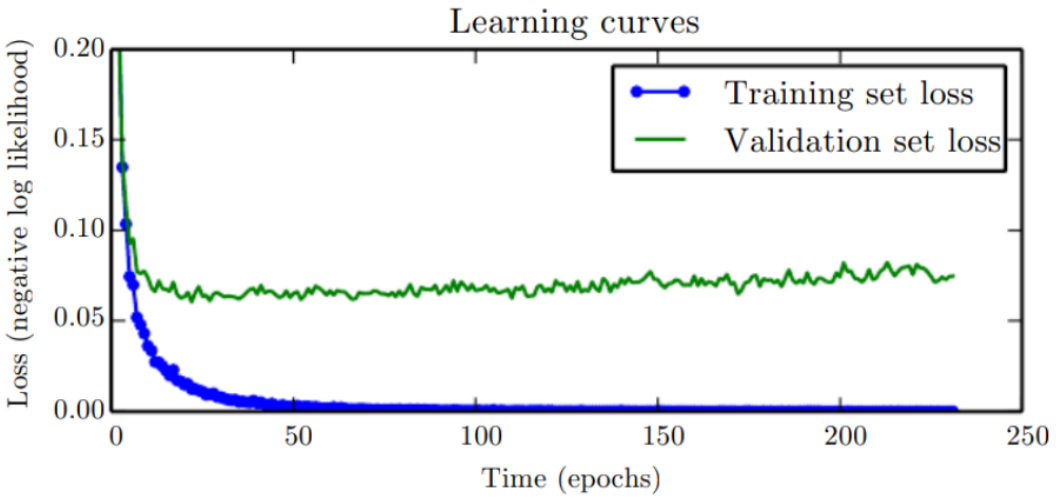
Randomly Flipping Digits Horizontally. This operation doesn't make sense for the MNIST dataset because digits have a specific orientation, and flipping them horizontally would result in incorrect representations (Flipping a digit "6" horizontally would make it look like a "9," leading to potential mislabeling).

**Early Stopping.** Another very popular technique in deep learning is early stopping (Deep Learning Book, section 7.8).

*1. How do the given loss curves (Deep Learning Book, figure 7.3) relate to early stopping?*

We implement the early stopping when the validation loss does not decrease or increase over a certain amount of time by stopping the training. If we take a look at the curve we see that after approximately 25 epochs we don't make any progress; it means we can stop training.

*2. Why is it a regularization technique?*




Early stopping regularizes the training process by preventing the model from continuing to learn the noise or specific patterns in the training data that may not generalize well. It encourages the model to generalize better to new, unseen data by stopping training when the model starts to compromise performance on the validation set.