**Name & Student No.:**

- Abdel Kader, Schehat & 10064822

- Tarbouch, Johny & 10033994

- Asfari, Mustafa & 10075209

1. Optimizers: Explain in your own words what problems are solved by momentum, learning rate, and learning rate decay

- **Momentum**: Helps escape local minima and faster convergence. By considering the average of past gradients, like a memory, resulting in a smooth vector which gets added.
- **Learning Rate**: Controls the step size during weight updates. High learning rate leads to fast learning but may overshoot and low learning rate is for fine tuning but is slow.
- **Learning Rate Decay**: Start with high learning rate and gradually reduce the learning rate to ensure fast learning and fine-tuning in later stages to avoid overshooting optimal values.

2. Dropout: Explain in your own words how dropout regularization is used during training.

- Randomly set some neurons to 0 during training, leads to the network not relying on certain features. Forcing the network to learn a variety of robust features and prevent overfitting because it acts like ensemble modelling.
- During testing, all neurons are active and weights are scaled appropriately to “average out” the randomness.

3. Convolutions: Given a 10x10 image and a stride of 2, what filter size is needed to end with a 5x5 feature map?

- N = 10
- Stride = 2

$$Output\ size = \frac{\mathrm{N} - \mathrm{F}}{\mathrm{stride}} - 1$$

$$5 = \frac{10 - \mathrm{F}}{2} - 1$$

$$F = 2$$