**Links of code which I had taken**

1. **EWC, GEM:**

<https://github.com/ruinanzhang/Rotated_MNIST_Continual_Learning>

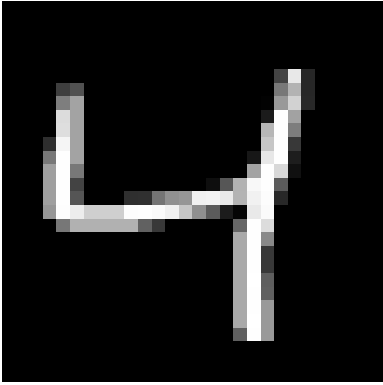
1. **Synaptic Intelligence:**

https://github.com/dchu1/AI\_P2\_cl

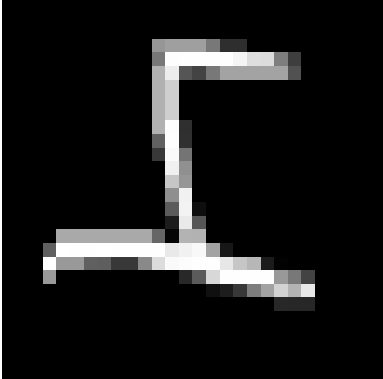
**Rotated MNIST dataset**

**We have performed our experiments on rotated MNIST dataset. For each rotation of each task, we rotate them for a random angle between 0 to 360**

**Original**

****

**Rotated**

****

1. **EWC**

**Link to code notebook:**

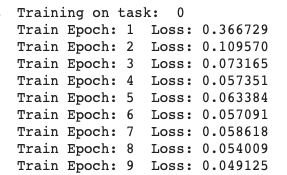
<https://colab.research.google.com/drive/1mL5NF8b1zkWlKAzCBv3z47lCZODaY7cf>

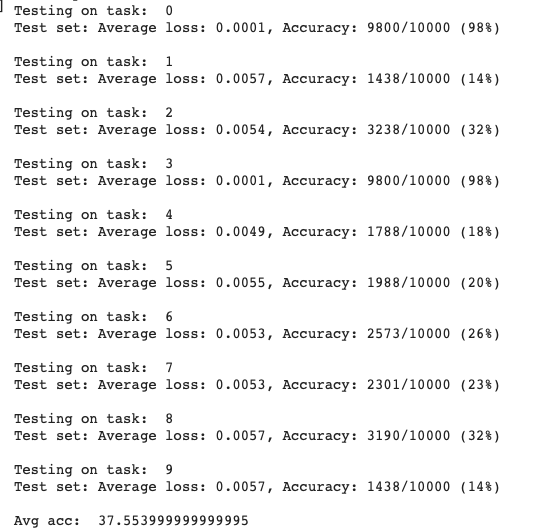
We decided to use MLP as the basic network architecture for our project. Our MLP model consists of four fully connected hidden layers -- with 512, 256,128,10 units in each hidden layer by themselves and use Rectified Linear Units(ReLUs) as activation function for each hidden layer. We also include a dropout layer (p=0.2) to prevent the overfitting of data.

We also tried different batch\_size ranging from 128 to 512, and finally decided to use batch size as 200 since this gave us the highest accuracy among all others

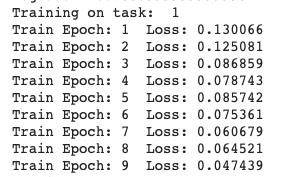
**As we trained our model on multiple tasks, we can see the accuracy for EWC improved from the 74% to 85.78%.**

**Training on Task 0 :**

****

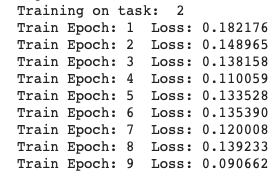
****

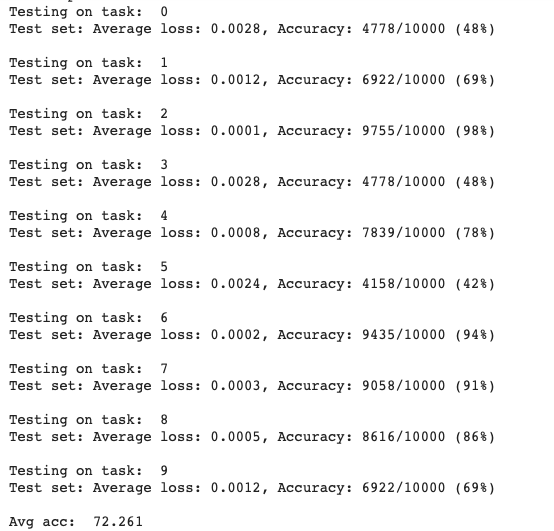
**Training on Task 1**

****

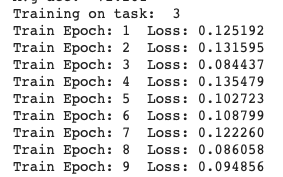
****

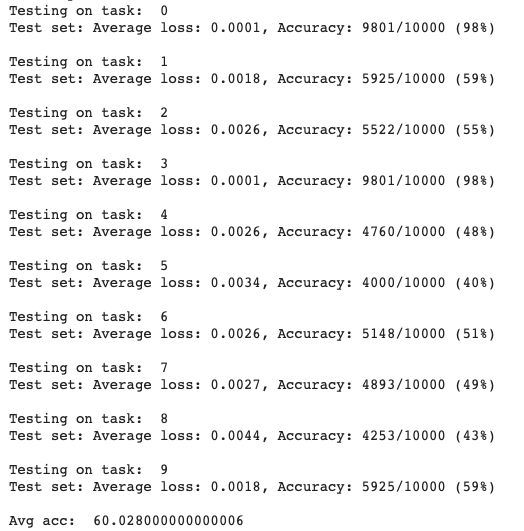
**Training on task 2**

****

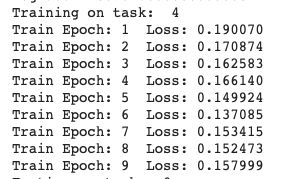
****

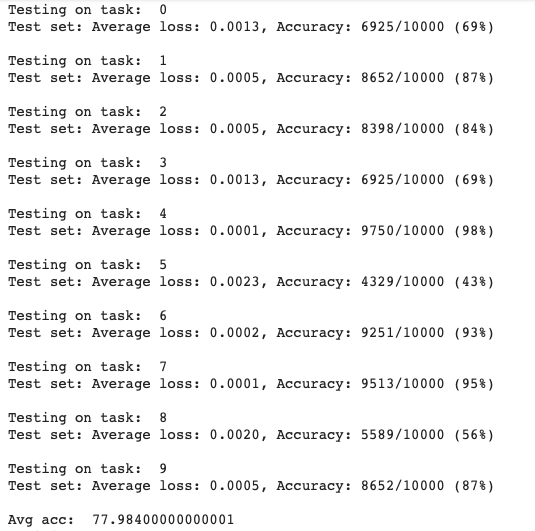
**Training on Task 3**

****

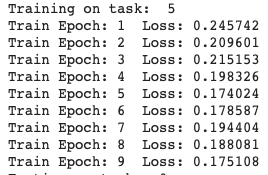
****

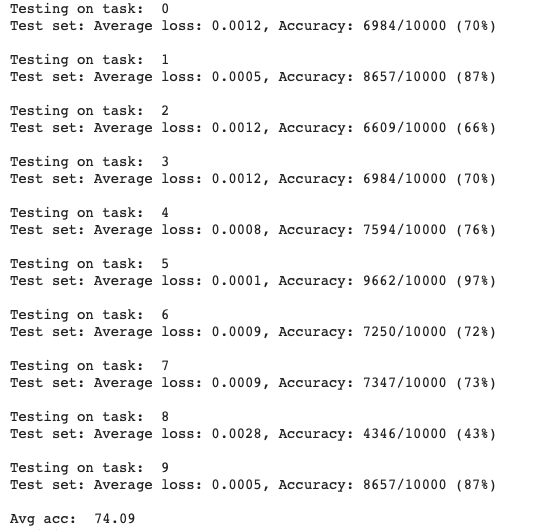
**Training on task 4**

****

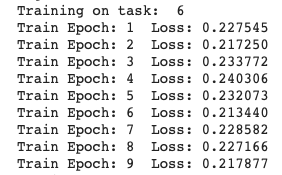
****

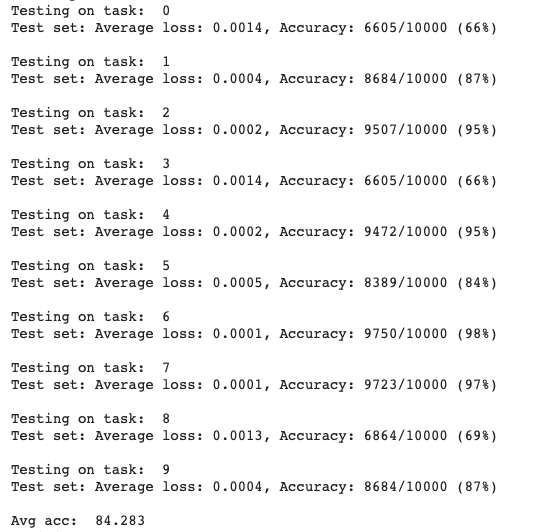
**Training on task 5**

****

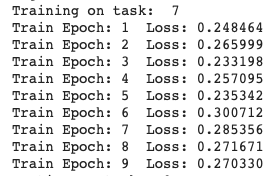
****

**Training on task 6**

****

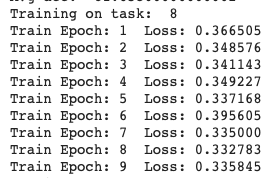
****

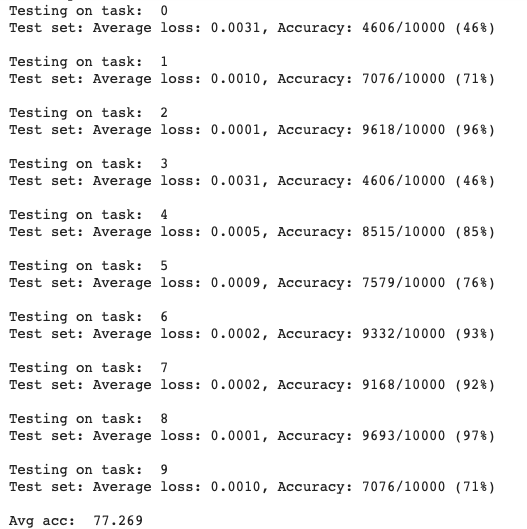
**Training on task 7**

****

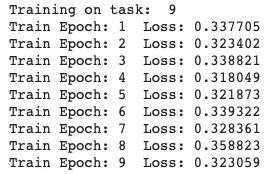
****

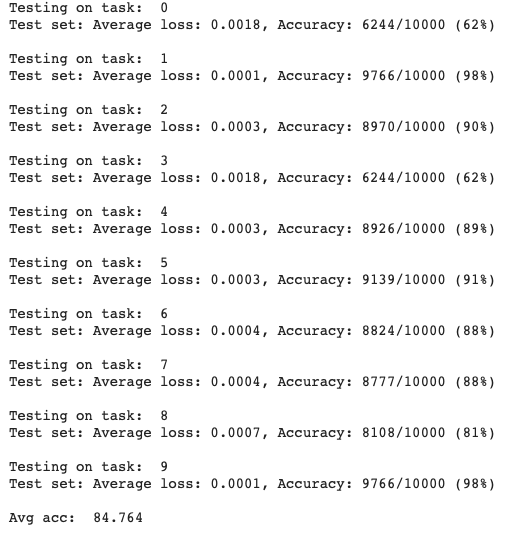
**Training on task 8**

****

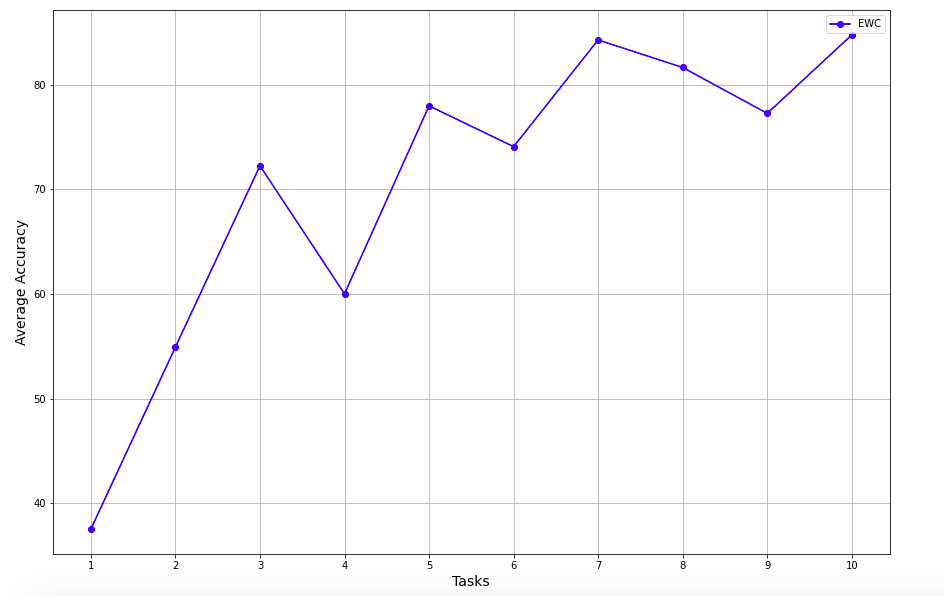
****

**Training on task 9**

****

****

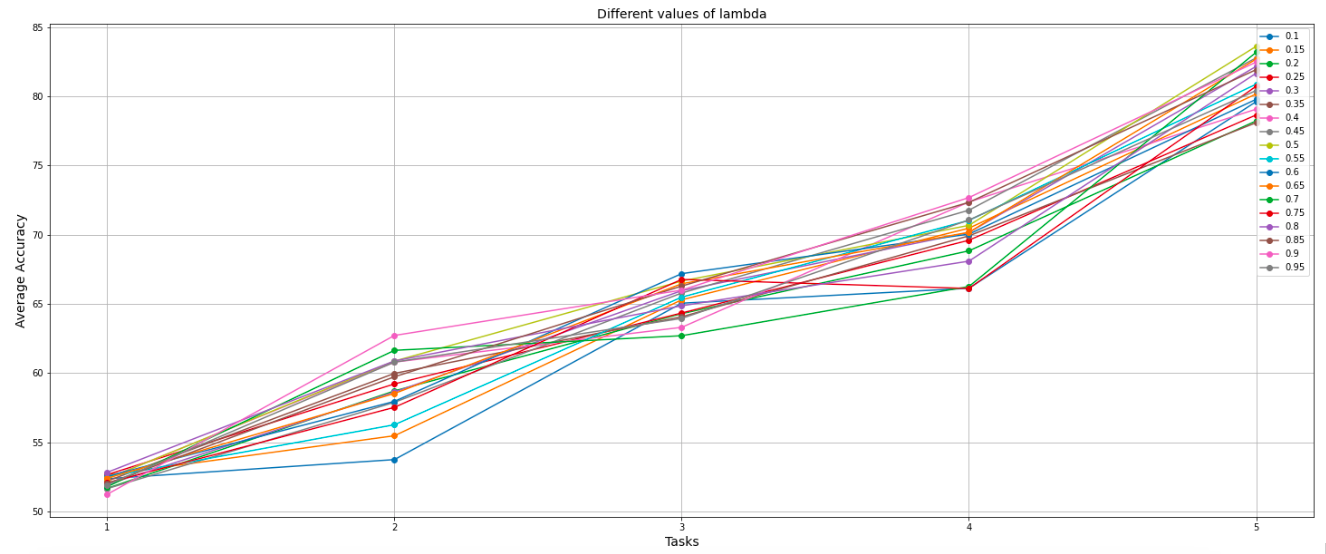
**Average Accuracy vs Number of Tasks**

****

**Exploring different values of lambda (**

In our loss function, (lambda) gives the importance of previous task’s parameters when compared to new task

Here we try to find the optimum value of for maximum accuracy



1. **Synaptic Intelligence**

**Link to code notebook:**

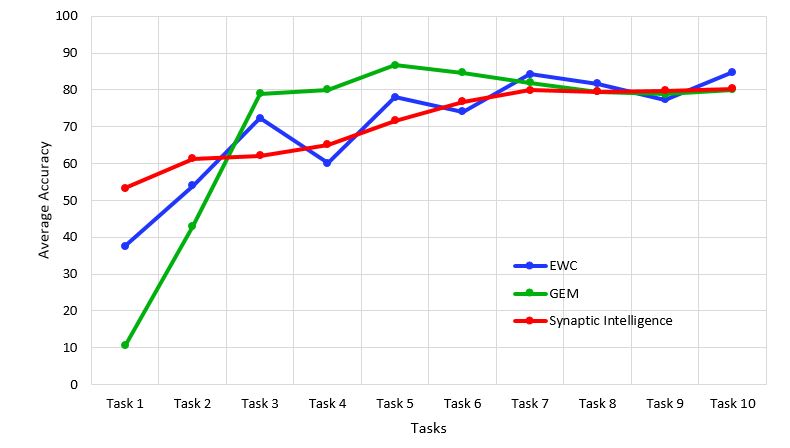
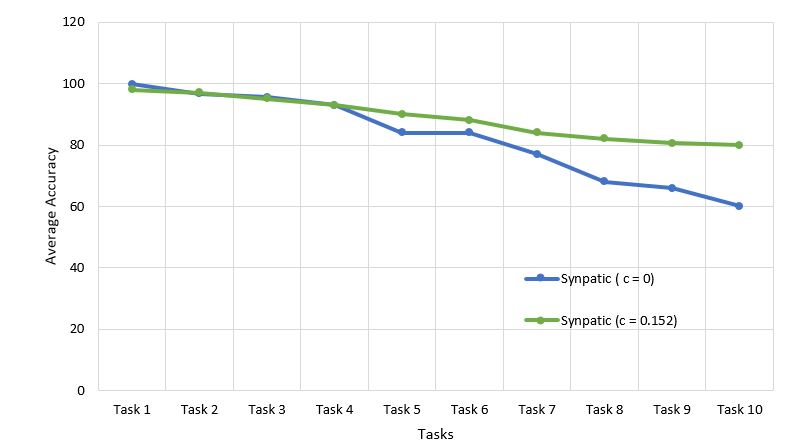
[**https://colab.research.google.com/drive/1HKpwwt9rgKkLleaFgpg2hV4LAsYmCpwL**](https://colab.research.google.com/drive/1HKpwwt9rgKkLleaFgpg2hV4LAsYmCpwL)

The model was trained on 10 tasks. After training we evaluated the model on the 10 test sets. For a control, we ran the experiment on the model using a *c* value of 0. This way the surrogate loss regularization term was always set to 0.

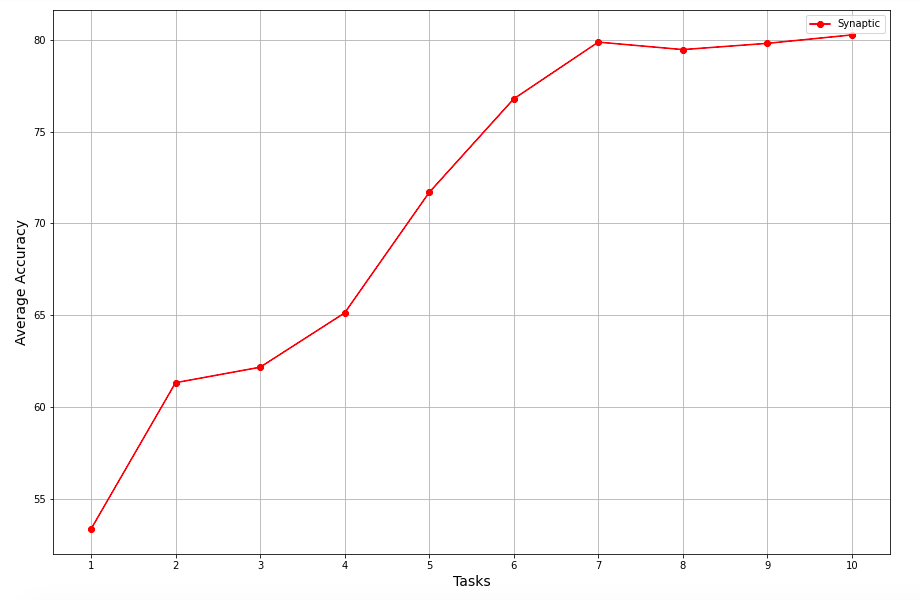
**Architecture:**

Our model consists of three fully connected hidden layers -- with 100, 100,100 units in each hidden layer by themselves and use Rectified Linear Units(ReLUs) as activation function for each hidden layer. We have used the adam optimizer and learning rate = 0.003.

Graph of average accuracy of the tasks seen vs the number of tasks for both control and experiment:

****

**Average Accuracy vs Number of Tasks**

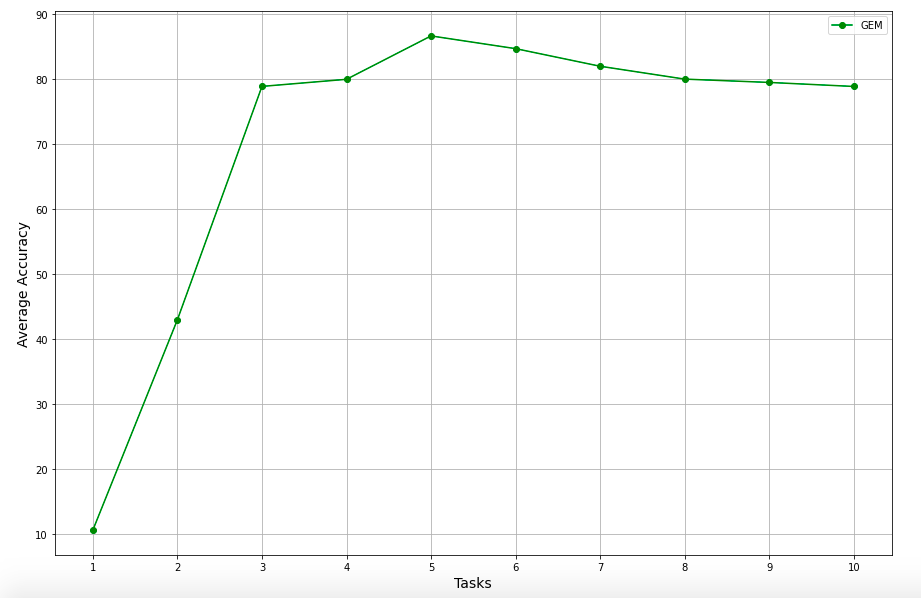
****

1. **GEM**

**Link to code folder:** <https://drive.google.com/drive/folders/1RCqAnQcEjt6p9l9D54H_AeJPY6eTk2pJ?usp=sharing>

When performing current task, GEM allows to access Episodic Memory where the subsets data of previous tasks being stored. We use integer task descriptors and focus on minimizing negative backward transfer (catastrophic forgetting) by the efficient use of episodic memory.

**Average Accuracy vs Number of Tasks**

****

**REFERENCES:**

Zenke, Friedemann, Ben Poole, and Surya Ganguli. "Continual learning through synaptic intelligence." In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 3987-3995. JMLR. org, 2017.

Kirkpatrick, James, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan et al. "Overcoming catastrophic forgetting in neural networks." *Proceedings of the national academy of sciences* 114, no. 13 (2017): 3521-3526.

Lopez-Paz, David, and Marc'Aurelio Ranzato. "Gradient episodic memory for continual learning." In *Advances in Neural Information Processing Systems*, pp. 6467-6476. 2017.

Li, Zhizhong, and Derek Hoiem. "Learning without forgetting." *IEEE transactions on pattern analysis and machine intelligence* 40, no. 12 (2017): 2935-2947.

Chen, Zhiyuan, and Bing Liu. "Lifelong machine learning." *Synthesis Lectures on Artificial Intelligence and Machine Learning* 12, no. 3 (2018): 1-207.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521, no. 7553 (2015): 436-444.

Ostapenko, Oleksiy, Mihai Puscas, Tassilo Klein, Patrick Jahnichen, and Moin Nabi. "Learning to remember: A synaptic plasticity driven framework for continual learning." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 11321-11329. 2019.

Fukushima, Kunihiko, and Sei Miyake. "Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition." In *Competition and cooperation in neural nets*, pp. 267-285. Springer, Berlin, Heidelberg, 1982.

Choy, Min Chee, Dipti Srinivasan, and Ruey Long Cheu. "Neural networks for continuous online learning and control." *IEEE Transactions on Neural Networks* 17, no. 6 (2006): 1511-1531.

Goodfellow, Ian J., Mehdi Mirza, Da Xiao, Aaron Courville, and Yoshua Bengio. "An empirical investigation of catastrophic forgetting in gradient-based neural networks." *arXiv preprint arXiv:1312.6211* (2013).

Redondo, Roger L., and Richard GM Morris. "Making memories last: the synaptic tagging and capture hypothesis." *Nature Reviews Neuroscience* 12, no. 1 (2011): 17-30.

Lahiri, Subhaneil, and Surya Ganguli. "A memory frontier for complex synapses." In *Advances in neural information processing systems*, pp. 1034-1042. 2013.

Ziegler, Lorric, Friedemann Zenke, David B. Kastner, and Wulfram Gerstner. "Synaptic consolidation: from synapses to behavioral modeling." *Journal of Neuroscience* 35, no. 3 (2015): 1319-1334.

Zenke, Friedemann, Everton J. Agnes, and Wulfram Gerstner. "Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks." *Nature communications* 6, no. 1 (2015): 1-13.

Fusi, Stefano, Patrick J. Drew, and Larry F. Abbott. "Cascade models of synaptically stored memories." *Neuron* 45, no. 4 (2005): 599-611.

Benna, Marcus K., and Stefano Fusi. "Computational principles of synaptic memory consolidation." *Nature neuroscience* 19, no. 12 (2016): 1697.