# Deep Learning Concepts

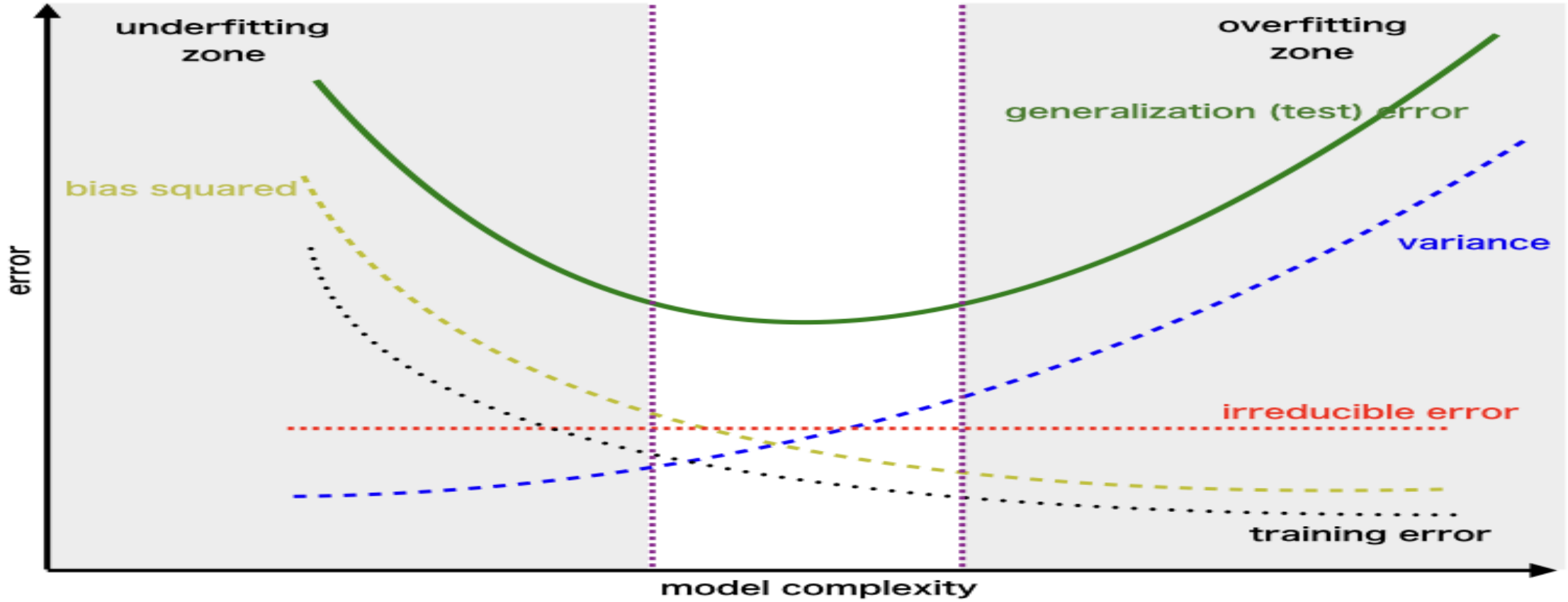
# Agenda

- Curse of dimensionality
- Bias - Variance trade-off
- Overfitting
- Underfitting
- Regularization techniques
- Importance of validation split

# Curse of Dimensionality

- The dimension of a dataset corresponds to the number of attributes/features that exist in a dataset.
- A dataset with a large number of attributes, generally of the order of hundred or more, is referred to as high dimensional data.
- The difficulties related to training machine learning models due to high dimensional data is referred to as '**Curse of Dimensionality**'.[1]

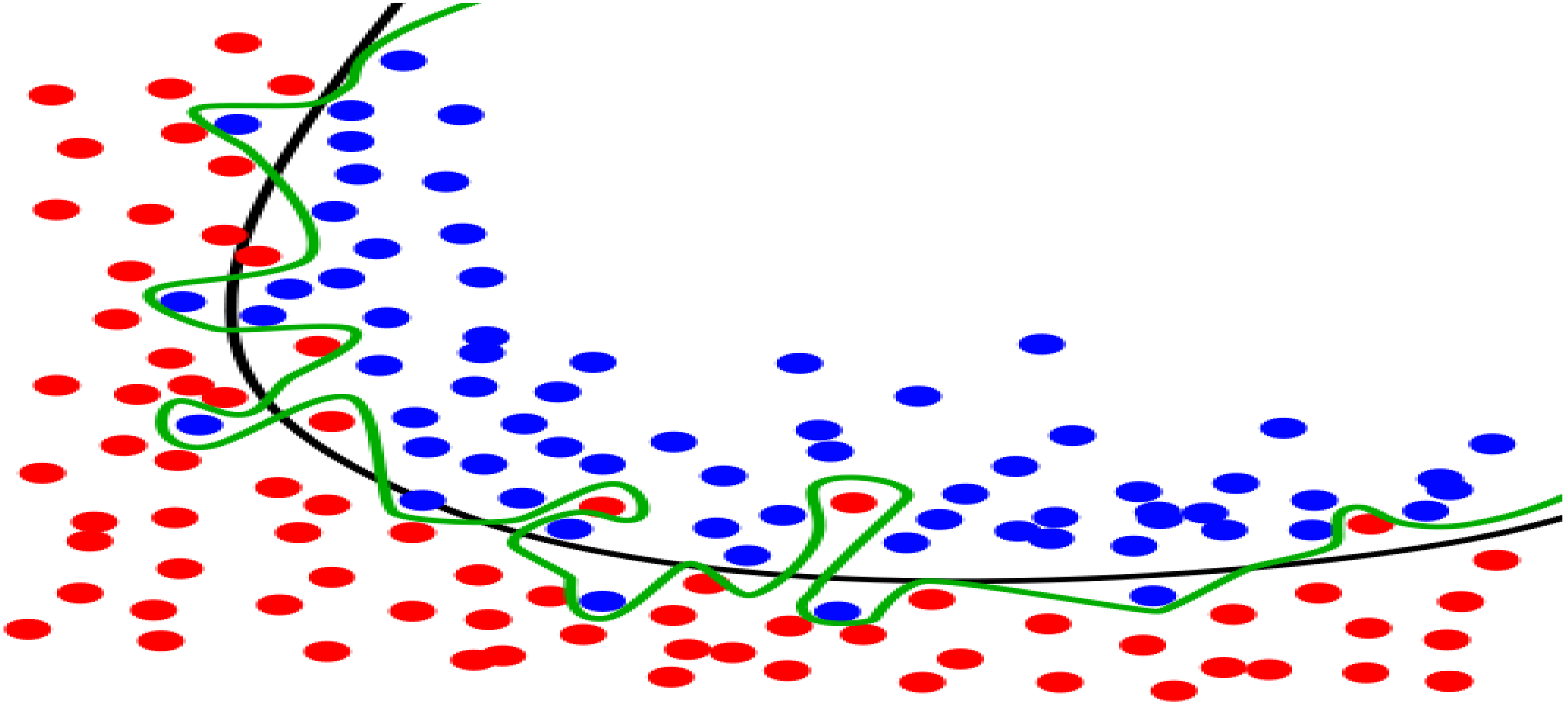
# Bias - Variance Tradeoff




Source - https://www.geeksforgeeks.org/ml-bias-variance-trade-off/ [2]

# Overfitting

- In statistics, **overfitting** is "the production of an analysis that corresponds too closely or exactly to a particular set of data and may therefore fail to fit additional data or predict future observations reliably".
- An **overfitted model** is a statistical model that contains more parameters than can be justified by the data.[3]



Source - Chabacano, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons [4]

# Underfitting

- Underfitting refers to a model that can neither model the training data nor generalize to new data.
- An underfit machine learning model is not a suitable model and will be obvious as it will have poor performance on the training data.[5]

# Why Regularization ?

Neural networks are complex in nature which makes them prone to overfitting. **Regularization** is a **technique** which makes slight modifications to the learning algorithm such that the model generalizes better.[6]

# Regularization Techniques

**L1 & L2 regularization**

update the general cost function by adding another term known as the regularization term.

*Cost function = Loss (say, binary cross entropy) + Regularization term*

Due to this the values of weight matrices decrease because it assumes that a neural network with smaller weight matrices leads to simpler models. Therefore, it will also reduce overfitting to quite an extent.[7]

## L2 regularizer

lambda is the regularization parameter. It is the hyperparameter whose value is optimized for better results. L2 regularization is also known as *weight decay* as it forces the weights to decay towards zero (but not exactly zero).

$$Cost\,function = Loss + \frac{\lambda}{2m} * \sum ||w||^2$$

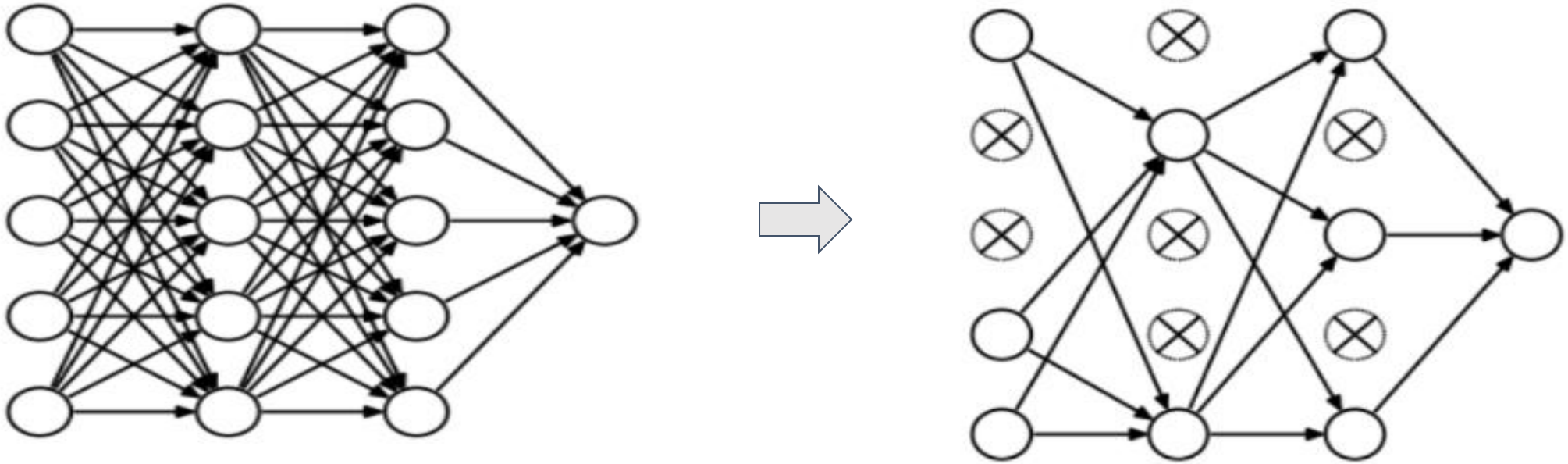
## L1 regularizer

In this, we penalize the absolute value of the weights. Unlike L2, the weights may be reduced to zero here. Hence, it is very useful when we are trying to compress our model. Otherwise, we usually prefer L2 over it.

$$Cost\,function = Loss + \frac{\lambda}{2m} * \sum \|w\|$$

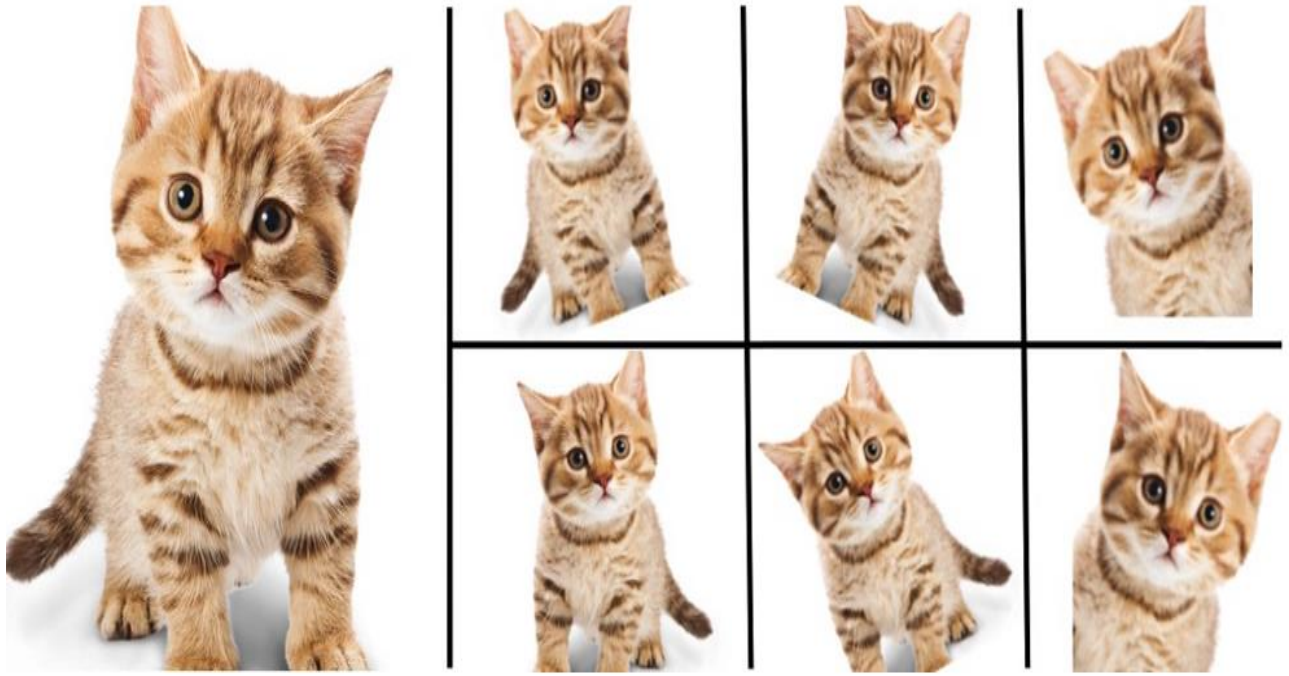
## Dropout

At every iteration, it randomly selects some nodes and removes them along with all of their incoming and outgoing connections. So each iteration has a different set of nodes and this results in a different set of outputs.[7]



# Data Augmentation

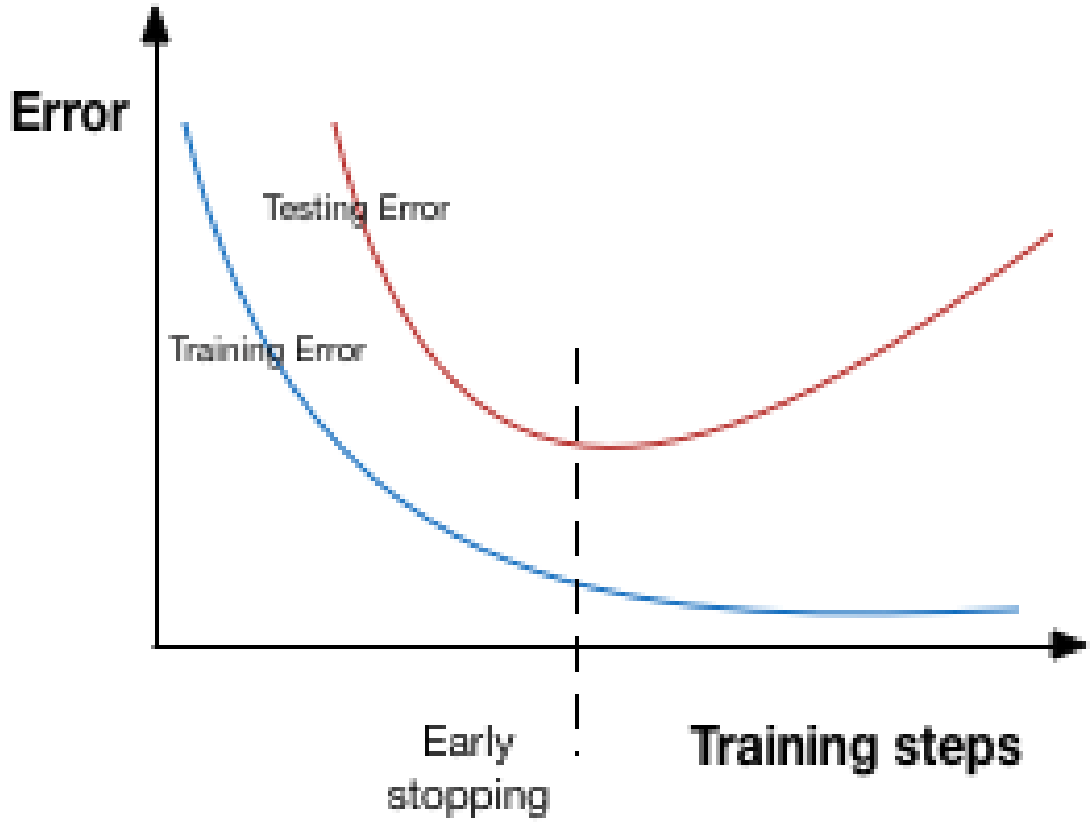
- Increase the size of the training data.
- Consider we are dealing with images. In this case, there are a few ways of increasing the size of the training data – rotating the image, flipping,scaling, shifting, etc. This technique is known as data augmentation.[7]




Source - https://www.kdnuggets.com/2018/05/data-augmentation-deep-learning-limited-data.html [8]

# Early Stopping

- One part of the training set as the validation set.
- When we see that the performance on the validation set is getting worse, we immediately stop the training on the model. This is known as early stopping.[7]




Source - https://www.analyticsvidhya.com/blog/2018/04/fundamentals-deep-learning-regularization-techniques/

# Importance of validation split

- A validation dataset is a sample of data held back from training your model that is used to give an estimate of model skill while tuning model's hyperparameters.
- It is different from the test dataset that is also held back from the training of the model, but is instead used to give an unbiased estimate of the skill of the final tuned model when comparing or selecting between final models.[8]

# REFERENCES


1. https://www.mygreatlearning.com/blog/understanding-curse-of-dimensionality/
2. https://www.geeksforgeeks.org/ml-bias-variance-trade-off/
3. https://en.wikipedia.org/wiki/Overfitting
4. Chabacano, CC BY-SA 4.0 , via Wikimedia Commons
5. https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning
6. https://www.analyticsvidhya.com/blog/2018/04/fundamentals-deep-learning-regularization
7. https://www.analyticsvidhya.com/blog/2018/04/fundamentals-deep-learning-regularization-techniques/
8. https://machinelearningmastery.com/difference-test-validation-datasets/

Microsoft

SAP

edunet foundation

# THANK YOU