NPTEL ONLINE CERTIFICATION COURSES

Course Name: Deep Learning

Faculty Name: Prof. P. K. Biswas
Department : E & ECE, IIT Kharagpur

Topic

Lecture 50: Training Tricks

CONCEPTS COVERED

Concepts Covered:

- Deep Neural Network
  - Normalization
  - Underfitting/Ovefitting
  - Regularization
  - Dropout
  - Early Stopping

Regularization
Early stopping

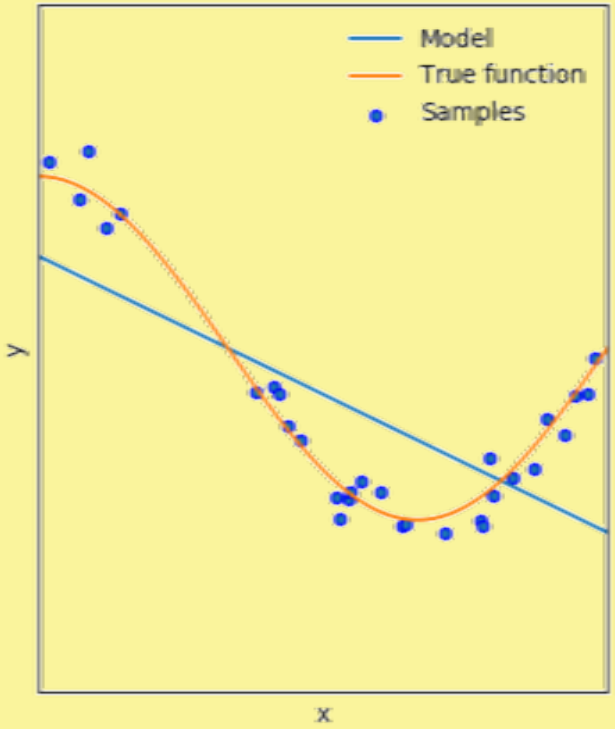
Overfitting/Underfitting

- Overfitting occurs when a statistical model or machine learning algorithm captures the noise of the data.
- Intuitively, overfitting occurs when the model or the algorithm fits the data too well.

- A statistical model or a machine learning algorithm is said to have underfitting when it cannot capture the underlying trend of the data.

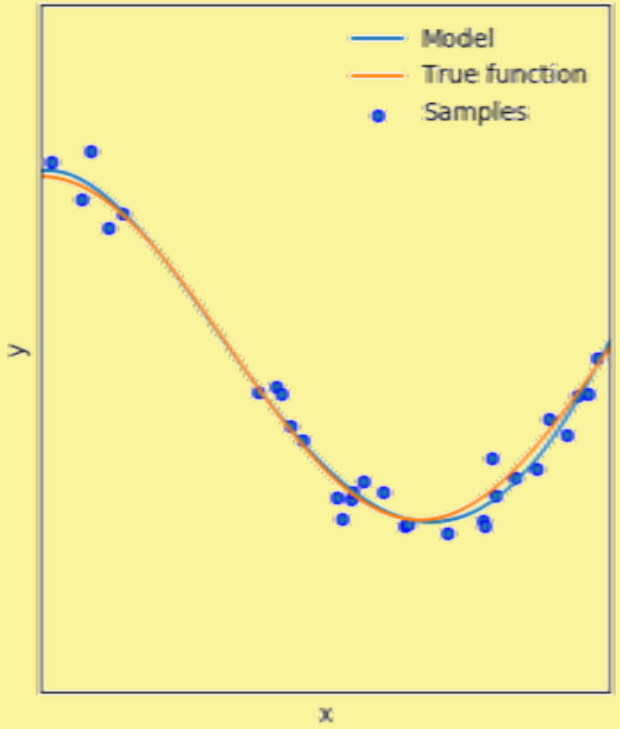
Overfitting/Underfitting: Regression

Degree: 1



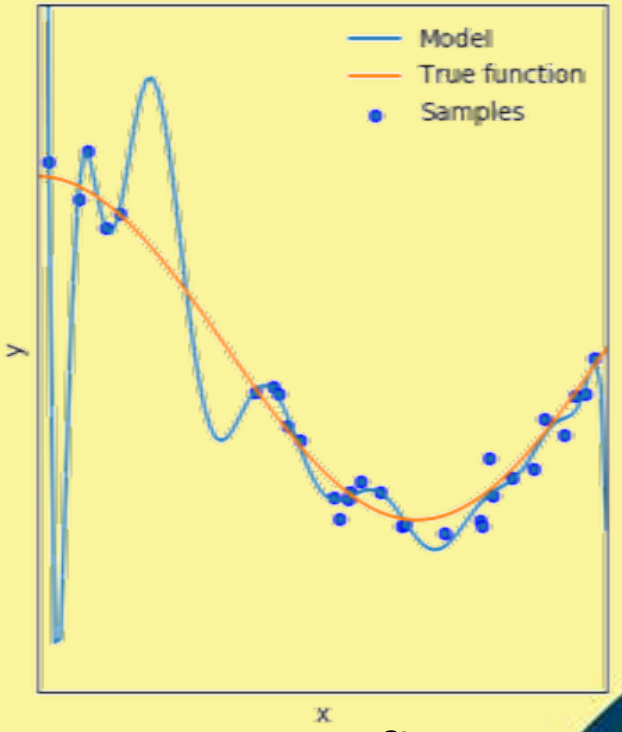
Underfit

Degree: 4



Perfectly fit

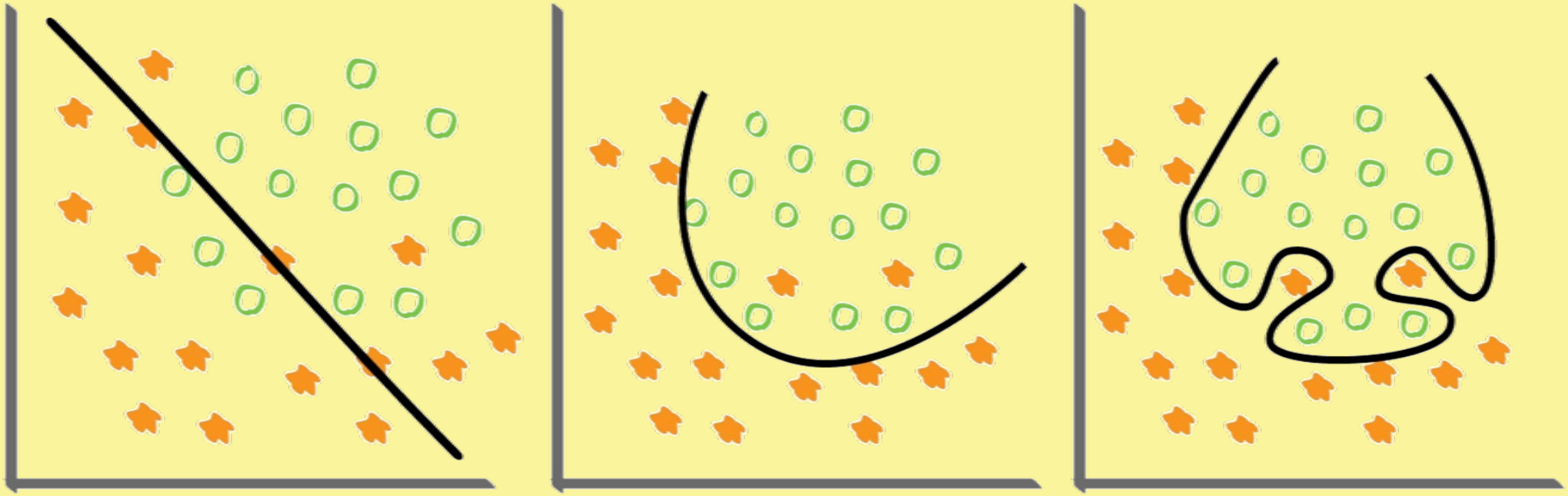
Degree: 15



Overfit

Image Source: Internet

Overfitting/Underfitting: Classification



Underfit **Perfectly fit** **Overfit**

Image Source: Internet

Regularization

- Regularization is a way to prevent overfitting.
- L1 and L2 are the most common types of regularization used in training deep models.
- General cost function with regularization for training is defined as: Cost function = Loss + Regularization term
- Due to this regularization term, the numerical values of weights decrease because it assumes that a neural network with smaller weights leads to simpler models.
- So this helps to reduce overfitting.

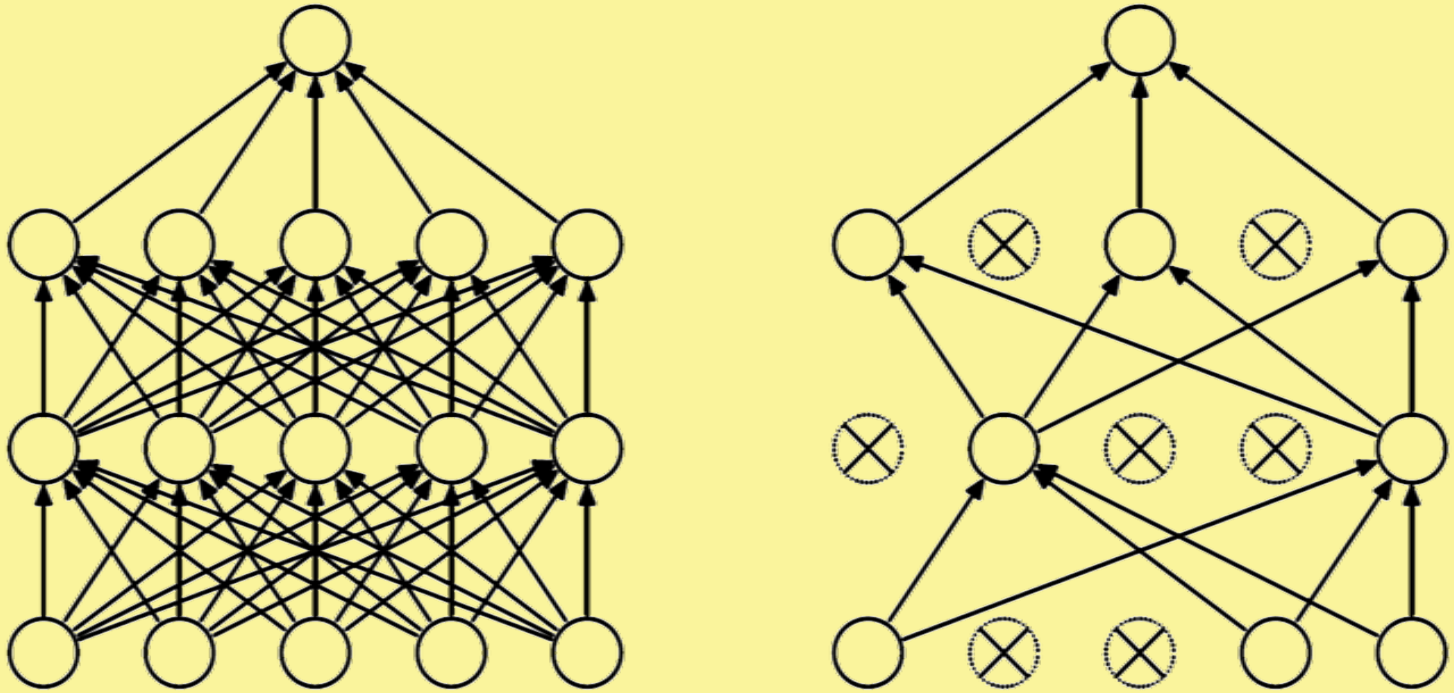
Regularization: L1 & L2

- L1 regularizer: Cost function = Loss + $\lambda \sum |w|$
  - It penalizes absolute value of weights
  - It can make some weights to zero. So useful for model compression.
  - λ is a regularization hyper parameter. Controls the relative weight.
- L2 regularizer: Cost function = Loss + $\lambda \sum ||w||^2$
  - It penalizes second norm of weights.
  - It is also termed as weight decay as it pushes the weights near to zero. But it does not make exactly zero always.

Data Augmentation

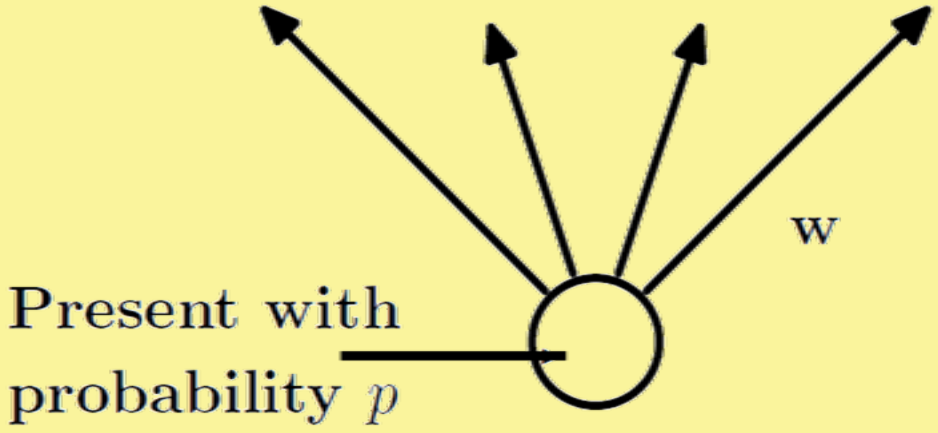
- Increasing the size of training data is a way to prevent overfitting.
- It is difficult and costly to increase the training data.
- Data augmentation is a way to create a different image from one image while keeping the context same.
- There are a few ways of augmenting training data– rotating, flipping, scaling, shifting, contrast enhancement, brightness control, etc.

Dropout

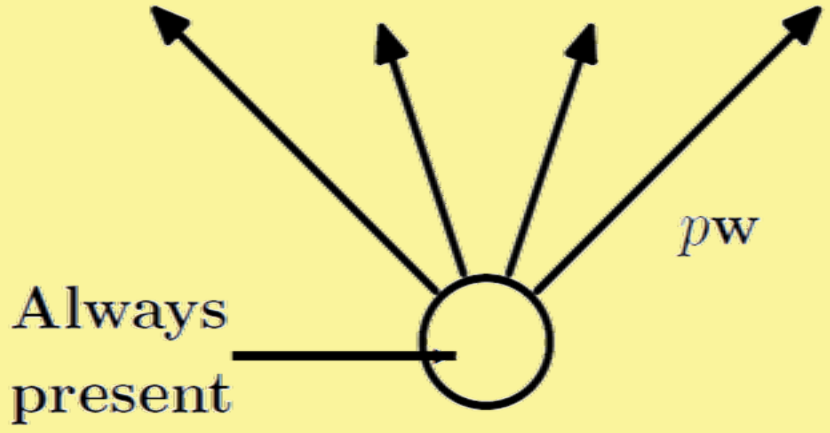


Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting." The Journal of Machine Learning Research 15.1 (2014)

Dropout



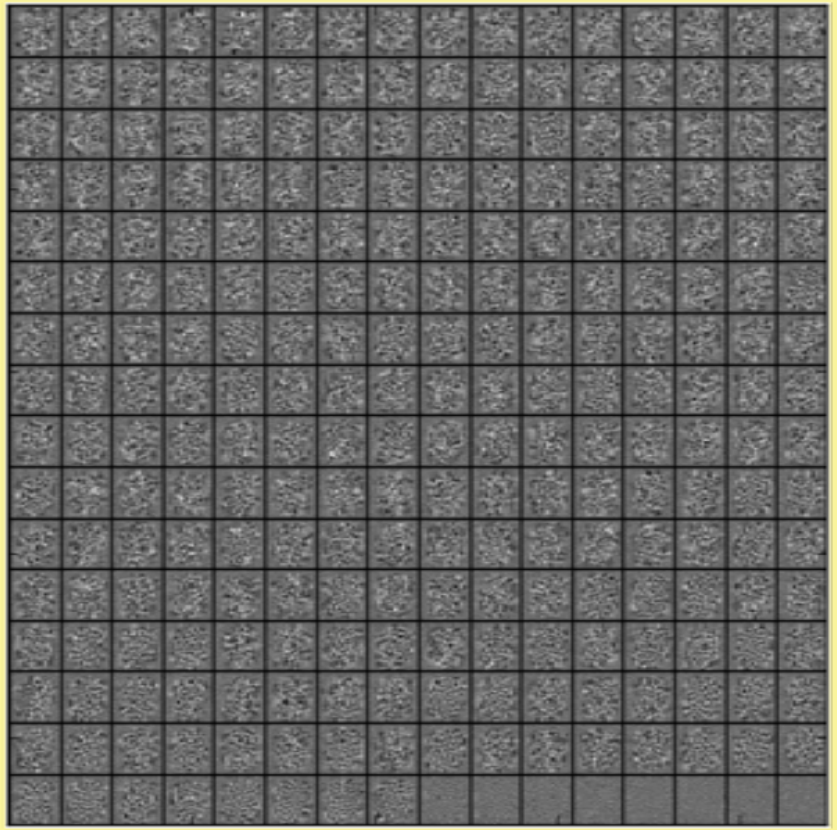
During Training



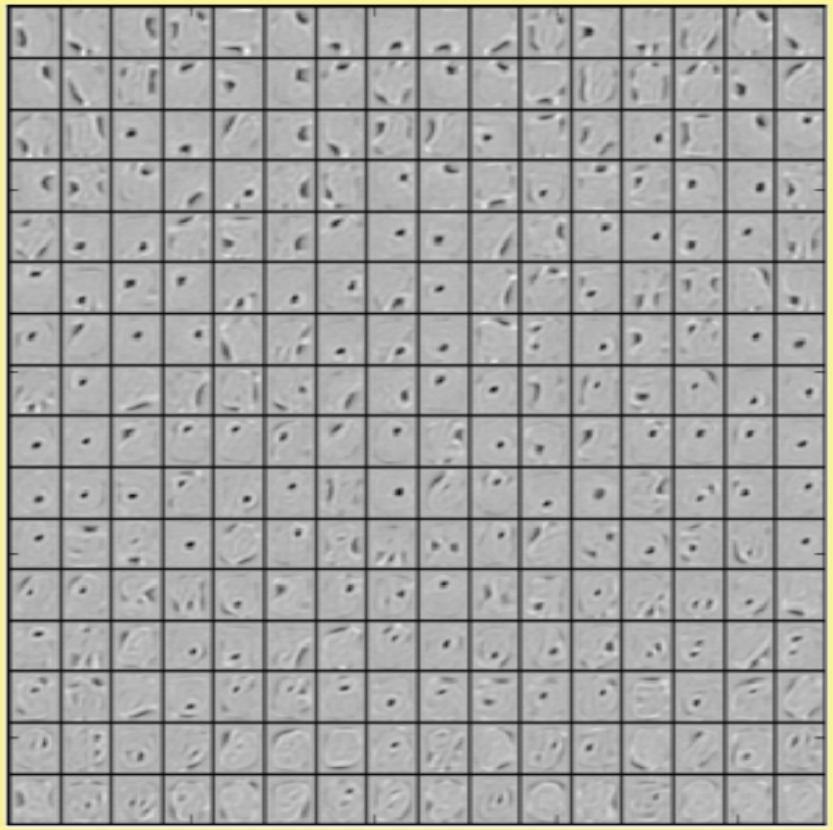
During Testing

Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting." The Journal of Machine Learning Research 15.1 (2014)

Dropout: Effect on learned features



Without dropout

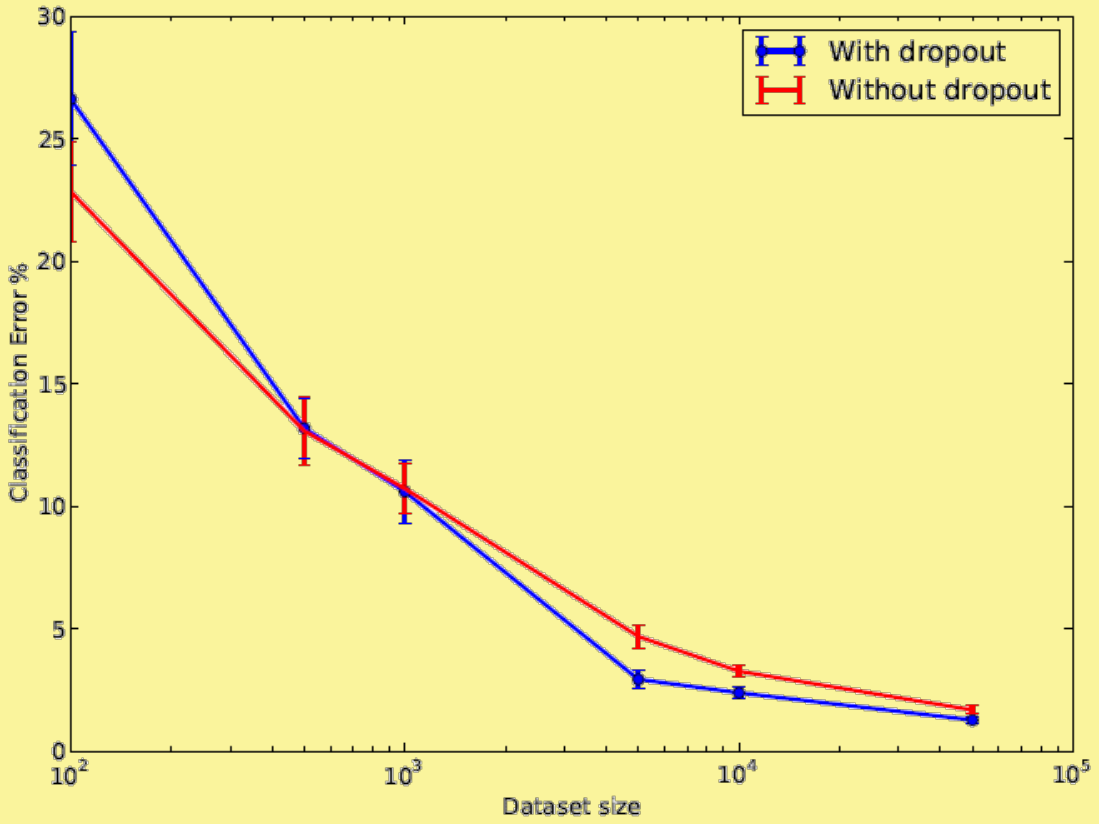


With dropout

Features learned by an autoencoder on MNIST with a single hidden layer of 256 rectified linear units with/ without dropout.

Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting." The Journal of Machine Learning Research 15.1 (2014)

Dropout: Effect on Data Size



- While model complexity is fixed, dropout does not generalize the model for very small amount of data
- As the size of the data set is increased, the gain from doing dropout increases up to a point and then declines.
- There is a sweet spot where amount of data is large enough.

Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting." The Journal of Machine Learning Research 15.1 (2014)

Early Stopping

- Hyperparameters need to be tuned for good performance while training neural networks.
- Number of iteration is a hyperparameter to be tuned. Lesser iteration may lead to underfit and more iteration may lead to overfit.
- Early stopping attempts to remove the need of manually setting this value.
- It can also be considered a type of regularization method.

Image Source: Internet

Early Stopping

- Hyperparameters need to be tuned for good performance while training neural networks.
- Number of iteration is a hyperparameter to be tuned. Lesser iteration may lead to underfit and more iteration may lead to overfit.
- Early stopping attempts to remove the need of manually setting this value.
- It can also be considered a type of regularization method.

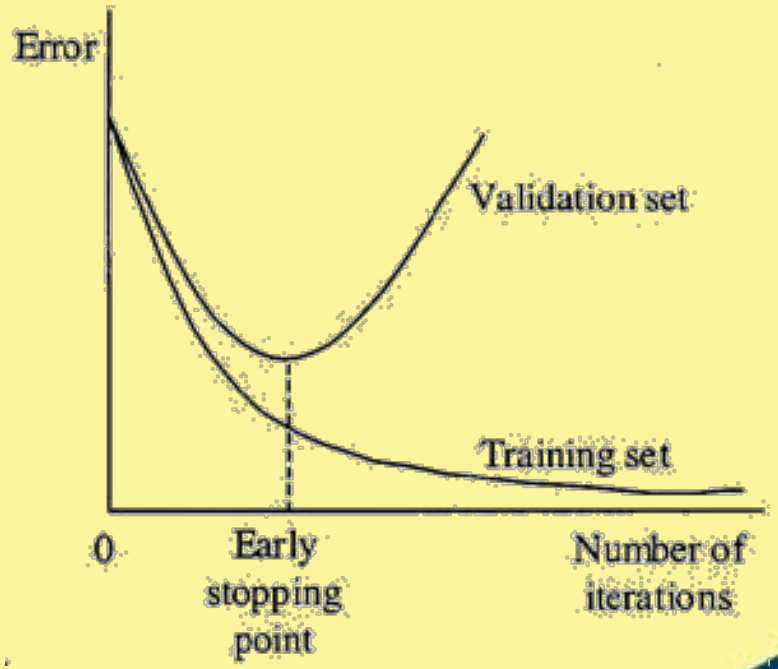


Image Source: Internet

Early Stopping

Early stopping algorithm is as follows:

- Split data into train, validation and test set
- After each training epoch:
  - Evaluate the model performance using validation data
  - Save the best model evaluated on validation data
- Use final model that has the best validation performance for testing.

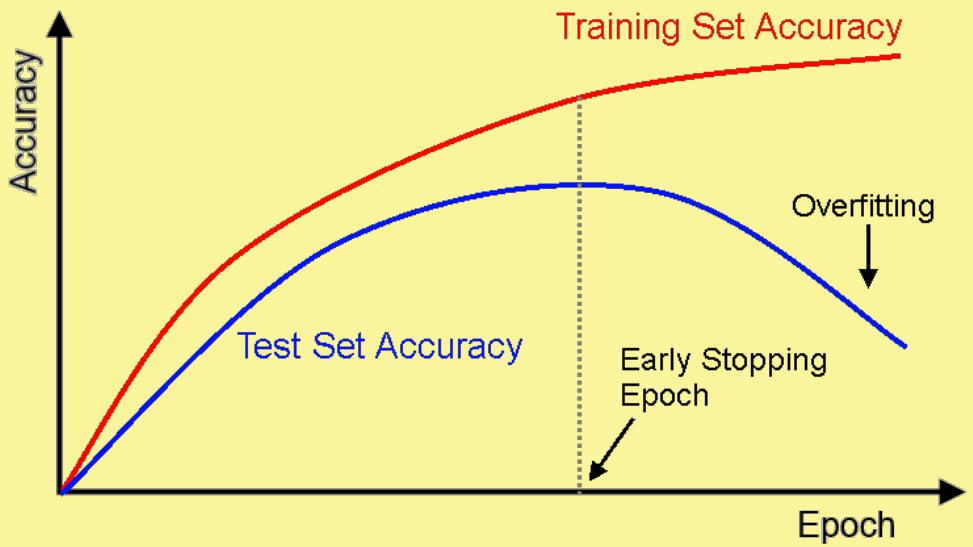


Image Source: Internet

NPTEL ONLINE CERTIFICATION COURSES

Thank you