Learning Curves

Learning curves plot models performance over experience or time.

Performance of machine learning models can be evaluated on training dataset to give an idea of how well the model is learning. It can be evaluated on validation dataset to get an idea on how well the model is generalizing.

- **Train Learning Curve:** Learning curve calculated on training dataset gives an idea on how well the model is learning.
- **Validation Learning Curve:** Learning curve calculated from validation set gives an idea of how well the model is generalizing.

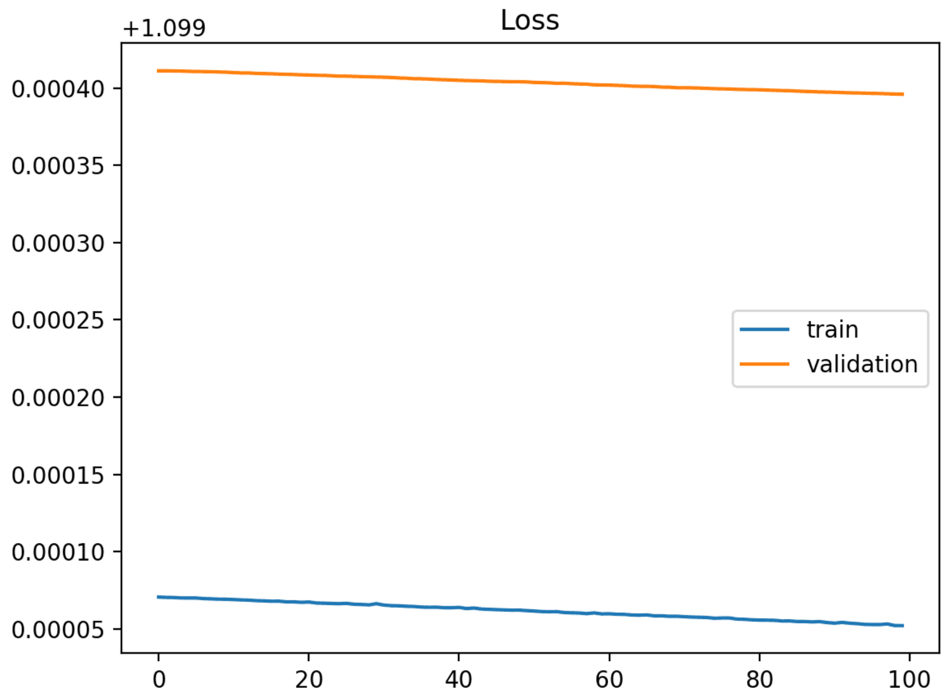
Normally below two kinds of dual curves are plotted for training and validation both.

- **Optmisation Learning Curves:** Learning curves calculated on the metric by which the parameters of the model are optimised. Eg. Loss
- **Performance Learning Curves:** Learning curves calculated on the metric by which the model will be evaluated or selected. Eg. Accuracy.

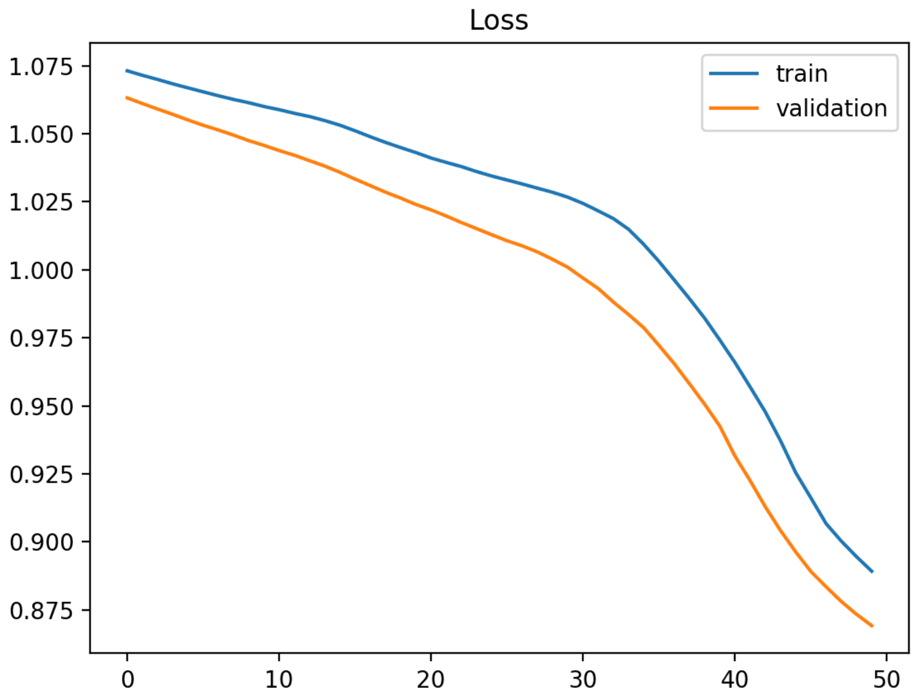
Diagnosing Model Behavior

Three common dynamics to observe in leaning curves:

- **Underfitting:** Underfitting occurs when model is not able to obtain sufficiently low error value on training set. This is normally plotted via training loss only.
    1. It may show flat line or noisy values of relatively high loss, indicating that model was unable to learn from training dataset. That means model is not having the capacity to learn from complicated dataset.



2. An underfit model may also be identified by a training loss that is deceasing and continues to decrease till end of the plot. This indicates that the model is capable of further learning and possible further improvement was halted prematurely.



A plot of learning curves shows underfitting if:

- The training loss remains flat regardless of training.
- The training loss continues to decrease until the end of training.

- **Overfitting**

  If we have too many feature the learned hypothesis may fit training data very well ($J(\theta) = 1/2m \sum_{i=1}^{m}(h_\theta x^{(i)} - y^{(i)})^2 = 0$), but fail to generalize new examples.

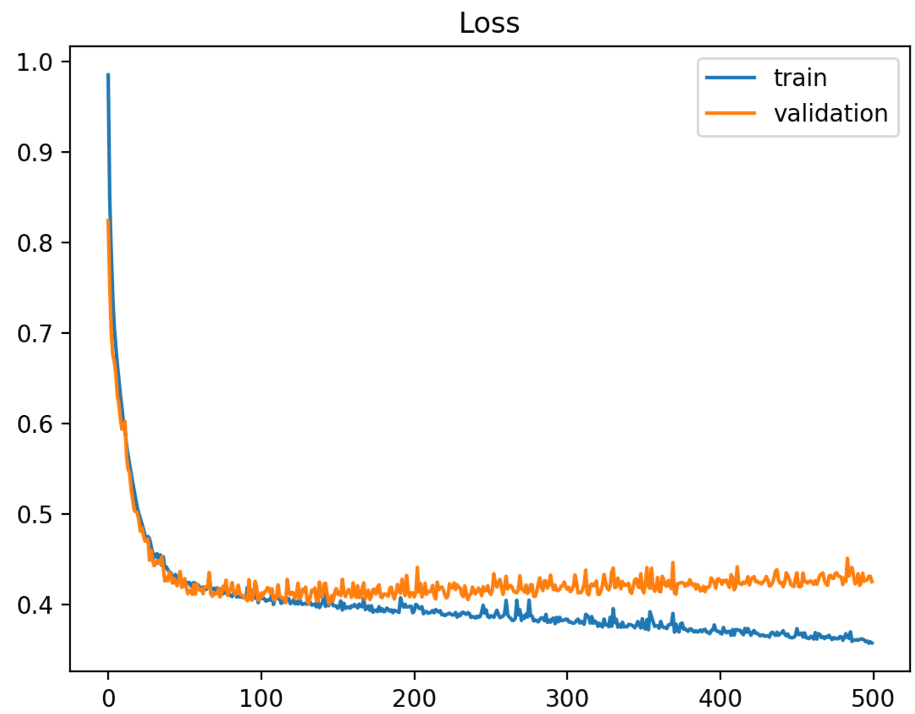
    Reduce the number of features.

    Selection of model

  Regularization - Keep all the features but reduce magnitude/values of the parameters θ_j. It works well when we have many features and each feature if contributing a bit in predicting y. The idea behind regularization is to keep smaller parameter values to **keep hypothesis simple and less prone to overfitting.**

  A plot looks like overfit if:

- The plot of training loss continues to decrease with experience.
- The plot of validation loss decreases to a point and begins increasing again.



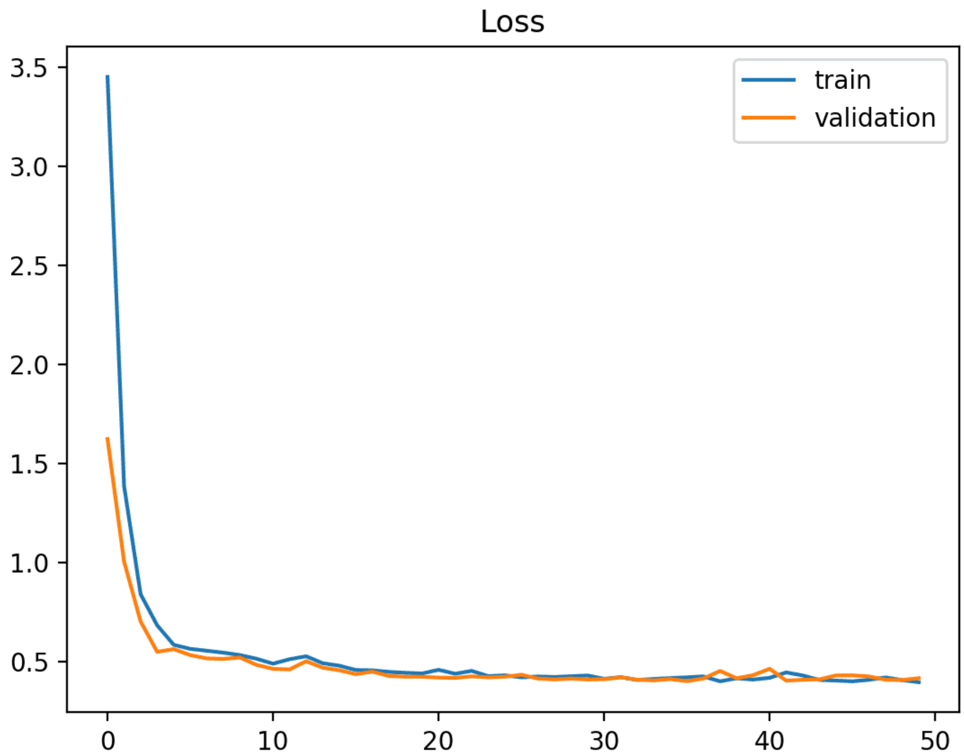
- **Good-fit**

  A good fit is the goal of the learning algorithm and exists between an overfit and underfit model.

  A good fit is identified by a training and validation loss that decreases to a point of stability with a minimal gap between the two final loss values.

  A plot looks like good-fit if:

  - The plot of training loss decreases to a point of stability.
  - The plot of validation loss decreases to a point of stability and has a small gap with the training l



More details of **Overfitting and Good to fit**