

Introduction to Learning Curves

Karel Horak

Brno University of Technology / Technische Universität Wien

horak@vut.cz

Slides modified, courtesy of Jason Brownlee

Overview

1. Learning curves
2. Diagnosing model behaviour
3. Diagnosing unrepresentative datasets

Learning Curves in Machine Learning

- Generally, a learning curve is a plot that shows time or experience on the x-axis and learning or improvement on the y-axis.
- **Learning curve** is line plot of learning (y-axis) over experience (x-axis).
- The metric used to evaluate learning could be maximizing, meaning that better scores (larger numbers) indicate more learning. An example would be classification accuracy.
- It is more common to use a score that is minimizing, such as loss or error whereby better scores (smaller numbers) indicate more learning and a value of 0.0 indicates that the training dataset was learned perfectly and no mistakes were made.

Learning Curves in Machine Learning

- During the training of a machine learning model, the current state of the model at each step of the training algorithm can be evaluated. It can be evaluated on the training dataset to give an idea of how well the model is “learning.” It can also be evaluated on a hold-out validation dataset that is not part of the training dataset. Evaluation on the validation dataset gives an idea of how well the model is “generalizing.”
- **Train learning curve** is learning curve calculated from the training dataset that gives an idea of how well the model is learning.
- **Validation learning curve** is learning curve calculated from a hold-out validation dataset that gives an idea of how well the model is generalizing.
- It is common to create dual learning curves for a machine learning model during training on both the training and validation datasets.

Learning Curves in Machine Learning

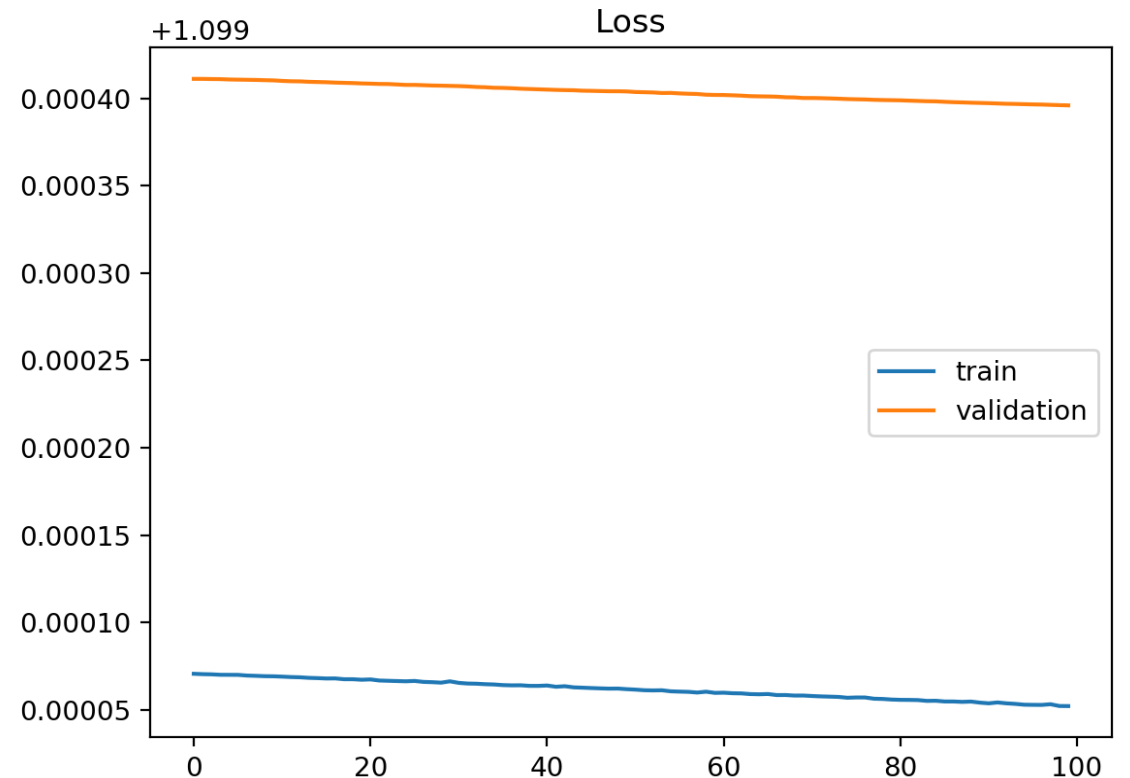
- In some cases, it is also common to create learning curves for multiple metrics, such as in the case of classification predictive modelling problems, where the model may be optimized according to cross-entropy loss and model performance is evaluated using classification accuracy.
- In this case, two plots are created, one for the learning curves of each metric, and each plot can show two learning curves, one for each of the train and validation datasets.
- **Optimization learning curves** are learning curves calculated on the metric by which the parameters of the model are being optimized, e.g. loss.
- **Performance learning curves** are learning curves calculated on the metric by which the model will be evaluated and selected, e.g. accuracy.

Diagnosing Model Behaviour

- The shape and dynamics of a learning curve can be used to diagnose the behaviour of a machine learning model.
- There are three common dynamics that you are likely to observe in learning curves; they are:
 1. Underfit.
 2. Overfit.
 3. Good Fit.
- Following examples will assume that we are looking at a minimizing metric, meaning that smaller relative scores on the y-axis indicate more or better learning.

Underfit Learning Curves

- Underfitting refers to a model that cannot learn the training dataset.
- An underfit model can be identified from the learning curve of the training loss only.
- It may show a flat line or noisy values of relatively high loss, indicating that the model was unable to learn the training dataset at all.
- An example of this is provided on the right and is common when the model does not have a suitable capacity for the complexity of the dataset.



Underfit Learning Curves

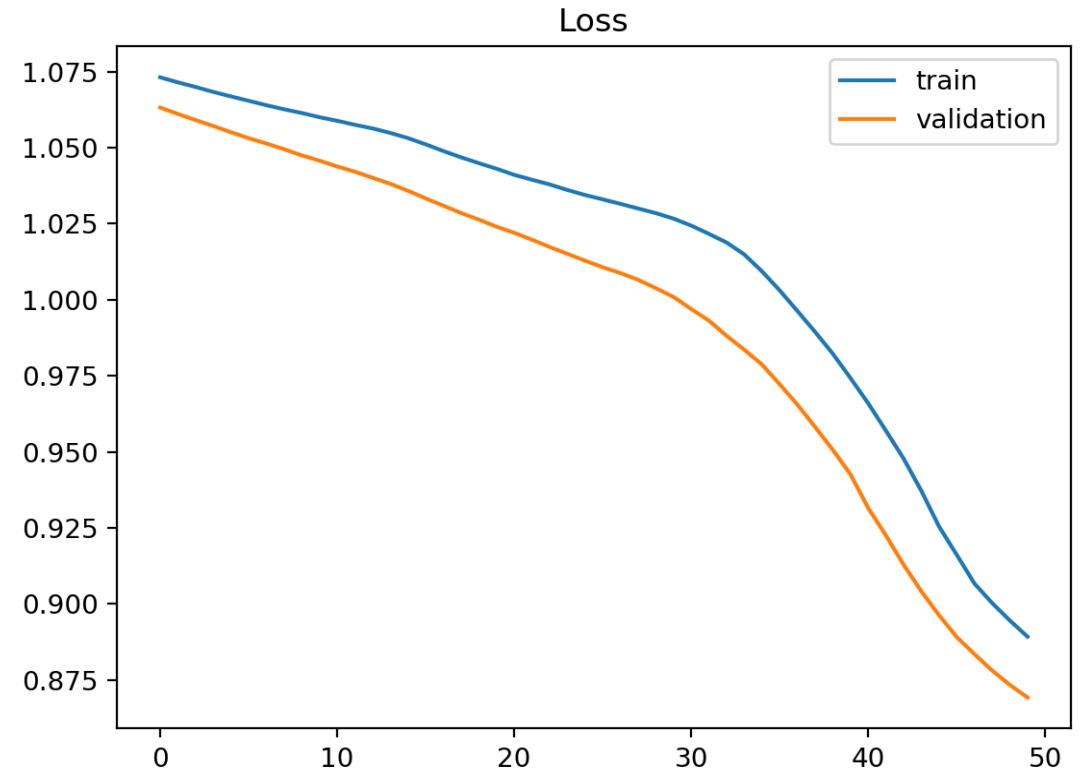
- An underfit model may also be identified by a training loss that is decreasing and continues to decrease at the end of the plot.
- This indicates that the model is capable of further learning and possible further improvements and that the training process was halted prematurely.

• A plot of learning curves shows underfitting if:

a) the training loss remains flat regardless of training

or

b) the training loss continues to decrease until the end of training.

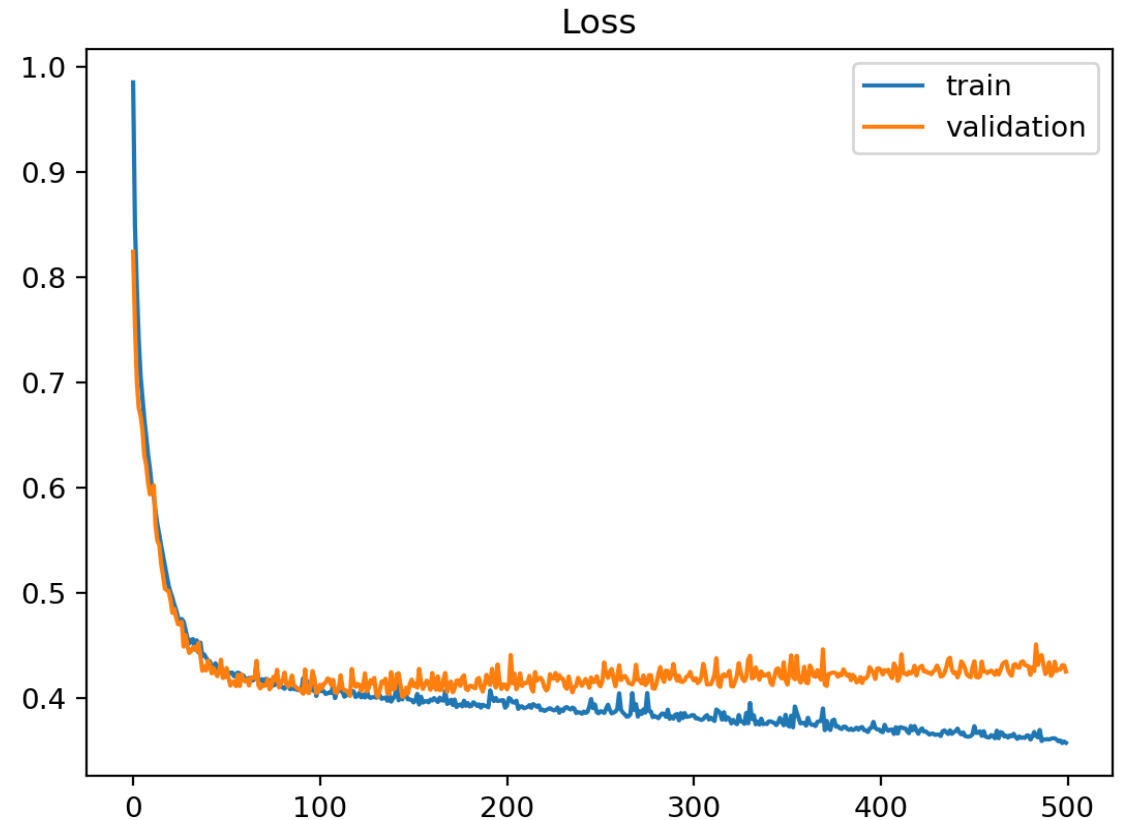


Overfit Learning Curves

- Overfitting refers to a model that has learned the training dataset too well, including the statistical noise or random fluctuations in the training dataset.
- The problem with overfitting, is that the more specialized the model becomes to training data, the less well it is able to generalize to new data, resulting in an increase in generalization error.
- This increase in generalization error can be measured by the performance of the model on the validation dataset.
- This often occurs if the model has more capacity than is required for the problem, and, in turn, too much flexibility. It can also occur if the model is trained for too long.

Overfit Learning Curves

- A plot of learning curves shows overfitting if:
 - a) the plot of training loss continues to decrease with experienceand
 - b) the plot of validation loss decreases to a point and begins increasing again.
- The inflection point in validation loss may be the point at which training could be halted as experience after that point shows the dynamics of overfitting.

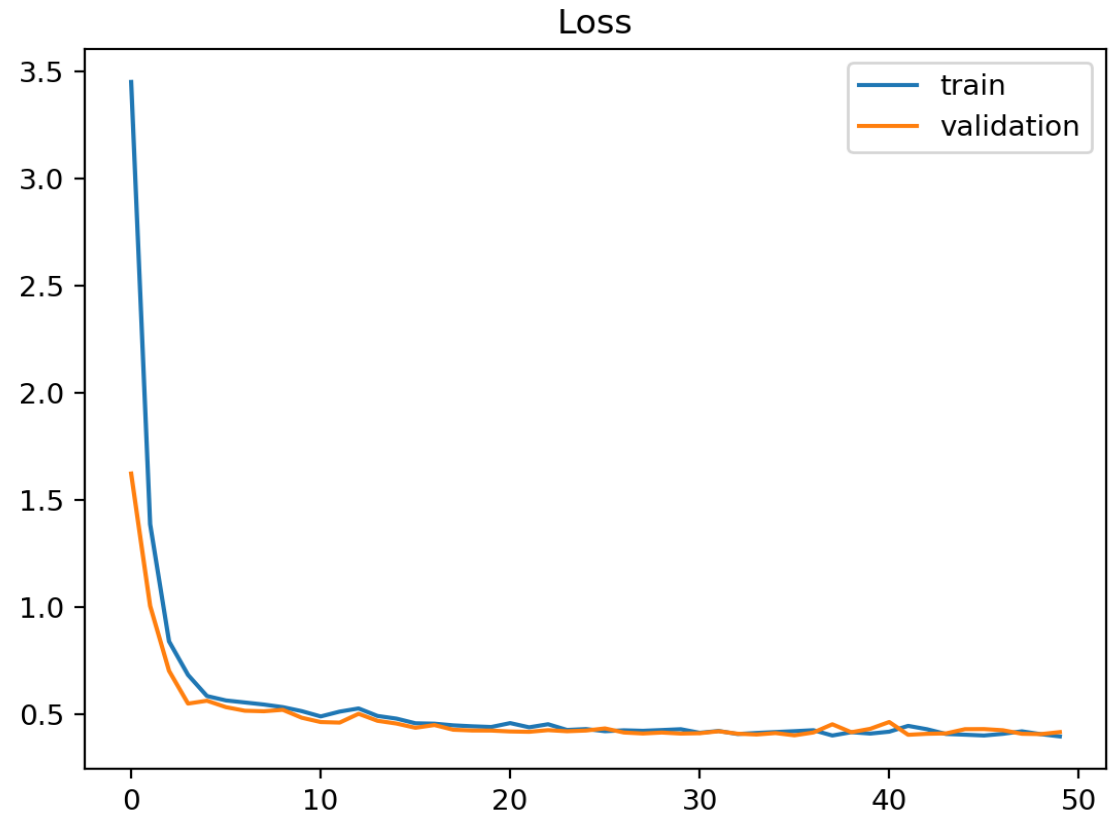


Good Fit Learning Curves

- 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.
- The loss of the model will almost always be lower on the training dataset than the validation dataset.
- This means that we should expect some gap between the train and validation loss learning curves.
- This gap is referred to as the “generalization gap”.

Good Fit Learning Curves

- A plot of learning curves shows a good fit if:
 - a) the plot of training loss decreases to a point of stability and
 - b) the plot of validation loss decreases to a point of stability and has a small gap with the training loss.
- Continued training of a good fit will likely lead to an overfit.

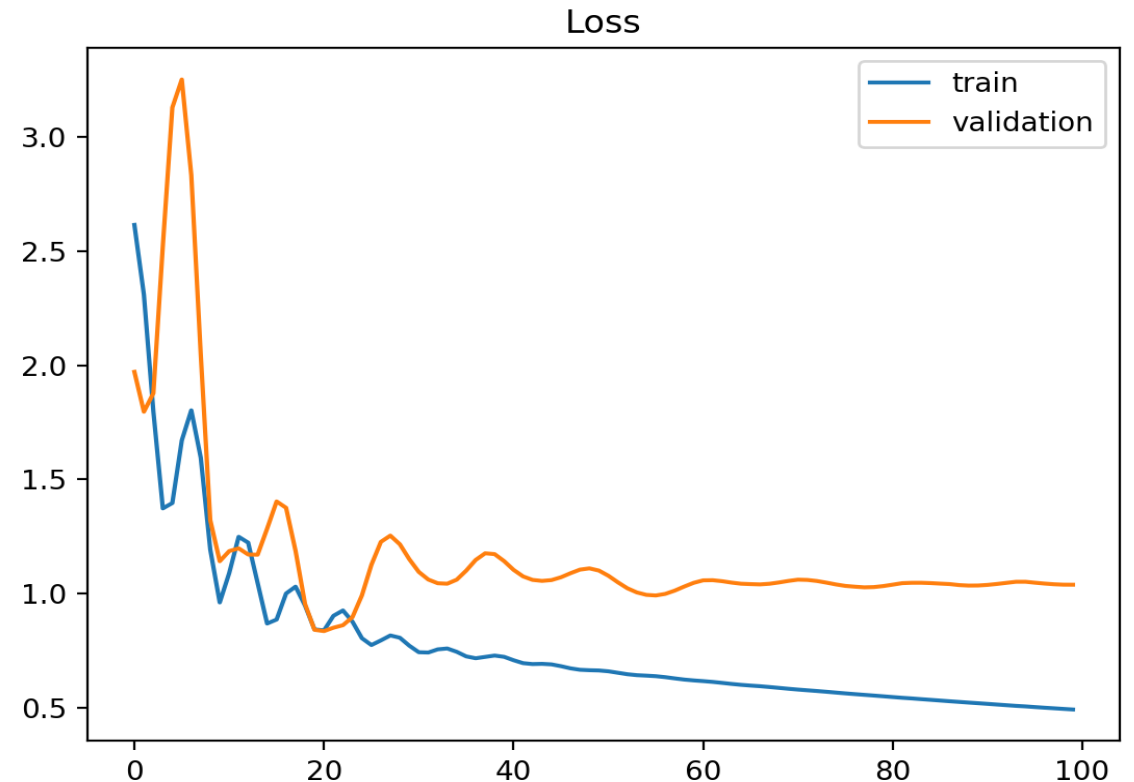


Diagnosing Unrepresentative Datasets

- Learning curves can also be used to diagnose properties of a dataset and whether it is relatively representative.
- An unrepresentative dataset means a dataset that may not capture the statistical characteristics relative to another dataset drawn from the same domain, such as between a train and a validation dataset.
- This can commonly occur if the number of samples in a dataset is too small, relative to another dataset.
- There are two common cases that could be observed; they are:
 1. training dataset is relatively unrepresentativeand / or
 2. validation dataset is relatively unrepresentative.

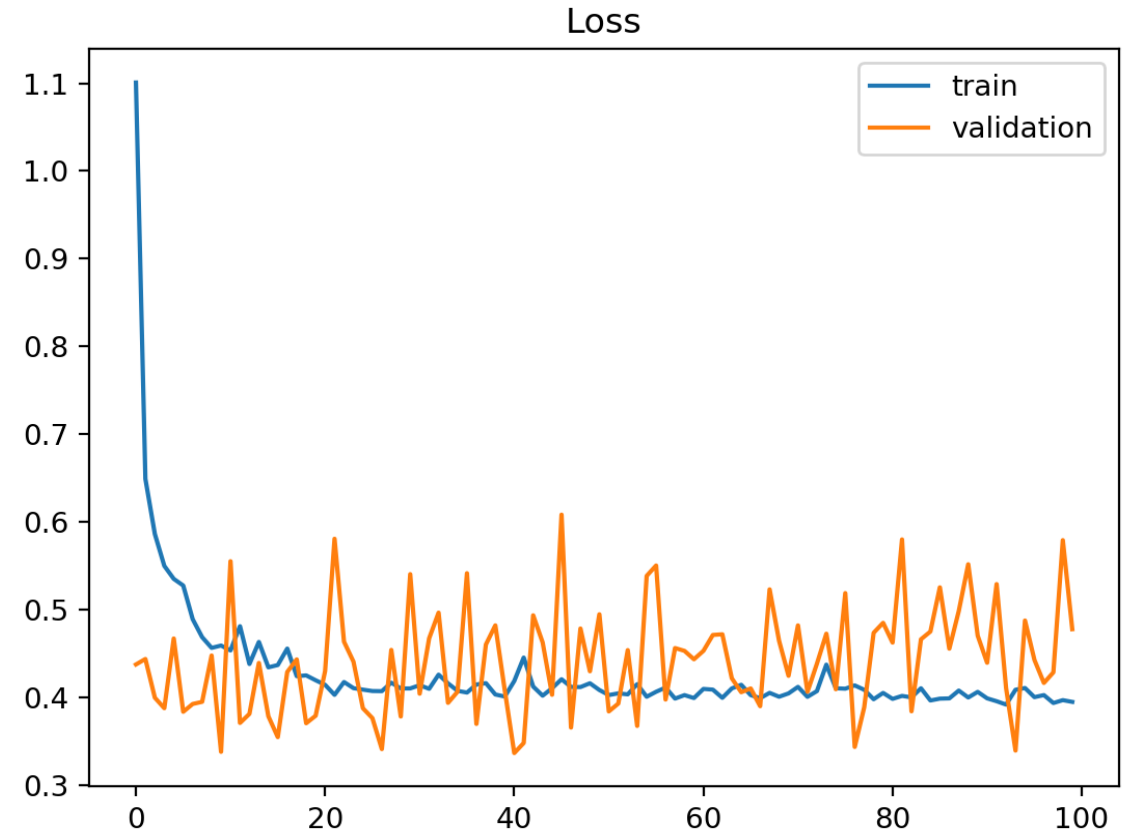
Unrepresentative Train Dataset

- An unrepresentative training dataset means that the training dataset does not provide sufficient information to learn the problem, relative to the validation dataset used to evaluate it.
- This may occur if the training dataset has too few examples as compared to the validation dataset.
- This situation can be identified by a learning curve for training loss that shows improvement and similarly a learning curve for validation loss that shows improvement, but a large gap remains between both curves.



Unrepresentative Validation Dataset

- An unrepresentative validation dataset means that the validation dataset does not provide sufficient information to evaluate the ability of the model to generalize.
- This may occur if the validation dataset has too few examples as compared to the training dataset.
- This case can be identified by a learning curve for training loss that looks like a good fit (or other fits) and a learning curve for validation loss that shows noisy movements around the training loss.



Unrepresentative Validation Dataset

- It may also be identified by a validation loss that is lower than the training loss.
- In this case, it indicates that the validation dataset may be easier for the model to predict than the training dataset.

