**Leaveoneout**

**Leave One Out Cross Validation** is just a special case of **K- Fold Cross** **Validation** where the number of folds = the number of samples in the dataset you want to run cross validation on.

Purpose of leave one out:

Due to the high number of test sets (which is the same as the number of samples) this cross-validation method can be very costly. For large datasets one should favor [**KFold**](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html#sklearn.model_selection.KFold), [**ShuffleSplit**](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.ShuffleSplit.html#sklearn.model_selection.ShuffleSplit) or [**StratifiedKFold**](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html#sklearn.model_selection.StratifiedKFold).

**Linear Discriminant Analysis** or LDA is a dimensionality reduction technique. It is used as a pre-processing step in [Machine Learning](https://www.knowledgehut.com/blog/data-science/what-is-machine-learning) .The goal of LDA is to project the features in higher dimensional space onto a lower-dimensional space in order to avoid the curse of dimensionality and also reduce resources and dimensional costs.

**TensorFlow**

The primary software tool of deep learning is TensorFlow. It is an open source artificial intelligence library, using data flow graphs to build models. It allows developers to create large-scale neural networks with many layers. TensorFlow is mainly used for: Classification, Perception, Understanding, Discovering, Prediction and Creation.

**Keras**

There are several differences between these two [frameworks](https://analyticsindiamag.com/top-5-recently-open-sourced-framework-for-developers/). Keras is a neural network library while TensorFlow is the [open-source library](https://analyticsindiamag.com/facebook-releases-open-source-library-for-3d-deep-learning-pytorch3d/) for a number of various tasks in machine learning. TensorFlow provides both high-level and low-level APIs while Keras provides only high-level APIs.

**MNIST**

The MNIST dataset is an acronym that stands for the **Modified National Institute of Standards and Technology dataset**.

**Lasso**

Trying to minimize the cost function, Lasso regression will automatically select those features that are useful, discarding the useless or redundant features. In Lasso regression, discarding a feature will make its coefficient equal to 0.

So, the idea of using Lasso regression for feature selection purposes is very simple: we fit a Lasso regression on a scaled version of our dataset and we consider only those features that have a coefficient different from 0

Batch Size

The batch size is a hyperparameter that defines the number of samples to work through before updating the internal model parameters.Think of a batch as a for-loop iterating over one or more samples and making predictions. At the end of the batch, the predictions are compared to the expected output variables and an error is calculated. From this error, the update algorithm is used to improve the model, e.g. move down along the error gradient.

**Stochastic Gradient Descent**. Batch Size = 1

**Stochastic Gradient Descent, or SGD for short, is an optimization algorithm used to train machine learning algorithms, most notably artificial neural networks used in deep learning.**

**The job of the algorithm is to find a set of internal model parameters that perform well against some performance measure such as logarithmic loss or mean squared error.**

**The algorithm is iterative. This means that the search process occurs over multiple discrete steps, each step hopefully slightly improving the model parameters.**

**Epochs**

**The number of epochs is a hyperparameter that defines the number times that the learning algorithm will work through the entire training dataset.One epoch means that each sample in the training dataset has had an opportunity to update the internal model parametersThe number of epochs is traditionally large, often hundreds or thousands, allowing the learning algorithm to run until the error from the model has been sufficiently minimized. You may see examples of the number of epochs in the literature and in tutorials set to 10, 100, 500, 1000, and larger.**