**Bias-Variance in Machine Learning:**

* **Bias:** A simple model that under-fits the data (learned too little). High bias can cause an algorithm to miss the relevant relations between features and target outputs
* **Variance:** A complex model that over-fits the data (learned too much). High variance can cause an algorithm to model the random noise in the training data, rather than the intended outputs

**Overfitting and its measures:**

* Algorithms that are too complex produce overfit models that memorize the noise instead of the signal. Variance refers to the algorithm's sensitivity to specific sets of training data. High variance algorithms will produce drastically different models depending on the training set
* Overfitting can be identified by checking validation metrics such as accuracy and loss. The validation metrics usually increase until a point where they start declining when the model is affected by overfitting
* The data can be separated into two different subsets. The training set represents a majority of the available data (about 80%). The test set represents a small portion of the data set (about 20%)
* By segmenting the dataset, we can examine the performance can be measured using the percentage of accuracy observed in both data sets to conclude on the presence of overfitting
* Some of the methods used to prevent overfitting include training with more data, ensembling, data augmentation, data simplification, feature selection, cross-validation, and regularization

**Regularization:**

* Regularization is a technique used to avoid this overfitting problem. The idea behind regularization is that models that overfit the data are complex models that have for example too many parameters
* Overfitting can also be controlled by increasing the size of the training dataset. Define a loss or cost function that describes how well the model fits the data. The goal is to find the model that minimizes this loss function
* The idea is to penalize this loss function by adding a complexity term that would give a bigger loss for more complex models. It significantly reduces the variance of the model, without a substantial increase in its bias

**Regularization Techniques:**

* There are few regularization techniques, namely Ridge Regression (L2 Norm), Lasso Regression (L1 Norm), Elastic Net Regression, Dropout, Data Augmentation, Early Stopping, and Ensembling
* In L1, the sum of the absolute values of the weights is imposed as a penalty while in L2, the sum of the squared values of weights is imposed as a penalty. They both differ in the way they assign a penalty to the coefficients
* L1 is usually preferred when we are interested in fitting a linear model with fewer variables. It is also useful when considering a categorical variable with many levels
* L2 is useful when there are a large number of variables with relatively smaller data samples, like in the case of genomic data
* Elastic-net is a compromise between the L1 and L2 regularisation that attempts to shrink and do a sparse selection simultaneously. Lambda is the regularization parameter whose value is optimized for better results

**References:**

* Jorge Leonel (Mar 31, 2019), Bias / Variance in Machine Learning, <https://medium.com/@jorgesleonel/bias-variance-in-machine-learning-656a1b58e1c9>
* Megha Mishra (May 26, 2018), REGULARIZATION: An important concept in Machine Learning, <https://towardsdatascience.com/regularization-an-important-concept-in-machine-learning-5891628907ea>
* Will Koehrsen (Jan 27, 2018), Overfitting vs. Underfitting: A Conceptual Explanation, <https://towardsdatascience.com/overfitting-vs-underfitting-a-conceptual-explanation-d94ee20ca7f9>
* Saurabh Singh (Oct 8, 2019), Regularisation Techniques in Machine Learning and Deep Learning, <https://medium.com/analytics-vidhya/regularisation-techniques-in-machine-learning-and-deep-learning-8102312e1ef3>
* Droupout is useful in neural networks and at every iteration it randomly selects some nodes and removes them along with all of their incoming and outgoing connections
* Data Augmentation is the simplest way to increase the size of the training data. This usually provides a big leap in improving the accuracy of the model
* Early stopping is a kind of cross-validation strategy where we keep 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
* Ensemble methods combine several machine learning techniques into one predictive model. There are a few different methods for ensembling, but the two most common are: Bagging and Boosting