* Learning is combination of Representation, Evaluation and Optimization.
  + Representation is the set legal functions that could fit a particular data set is hypothesis space. This is called representation of a learning algorithm.
  + Evaluation like accuracy, Confusion matrix, mean square error is the measure of how good is the classifier.
  + Optimization is the number of techniques used to reach the optimal solution. Efficiency of algorithm is affected by the optimization technique used.
* How good a classifier is, depends on how well it generalizes on the unseen data.
  + A classifier may perform very well on the training set. But, the function might be overfitting the data. It is considered a good classifier if it performs in a similar manner (low error) on the test data as well.
  + Cross validation could help resolve this problem.
* In machine learning, data alone is not enough to improve the classifier.
  + If the learning algorithm is suffering from high bias problem, it is very unlikely to improve the algorithm by collecting more data.
* Overfitting is caused when the classifier fails to generalize and fits the training data with almost 100% accuracy.
  + This is generally caused due to high variance when very high degree polynomial function is used for algorithm.
  + Similarly, a very low degree classifier causes high bias resulting in underfit problem.
  + Increasing the number of training set examples might help overcome the high variance problem but it may never help the high bias problem.
  + Decision tree can generally never suffer from high bias problem, though sometimes might end up with high variance(overfitting).
  + Adding a regularization term reduces the overfitting effect by penalizing the higher degree terms.
* Curse of dimensionality and blessing of non-uniformity deals with the number of features of the data set.
  + An algorithm suffering from high variance problem might never be improved by adding more number of features.
  + When the data is non-uniform and concentrated in some parts vs other, it helps in classifying it better. For example, KNN algorithm on digit data set works better since the space of digit image is much smaller than all possible images.
  + The intuition that we get in lower dimension might be disrupted in higher dimension.
* Feature engineering could help improve the learning algorithm. Creating new features that has higher co-relation with labels might help the classifier.
  + Some features alone might not be correlated with the target label, but when combined with other features, it might have a very good correlation with the label improve the learning algorithm.
* Enormous amount of that means more complex classifiers could be learnt, but practically simpler classifiers that gives good enough results are preferred over the complex ones because they take a lot of time to learn.
  + Fixed size learners like linear classifiers can never take advantage of huge amount data, while algorithms like decision tree which has a variable size could be improved with more data, but with computational cost trade-off.
* To improve the learning algorithm, training multiple models and taking out an average across them helps. It reduces the variance.
  + Techniques like bagging where random variations of dataset are generated by resampling, and the models generated from each one are voted or averaged over for the final result.
  + Similarly, techniques like boosting varies the weights of the training examples such that it punished those that got a bad result and boosts the ones that fetched correct results.
* Representable does not imply learnable. If a function is representable, does not necessarily mean that it is learnable.
  + We see complex functions with multiple minima which are representable but difficult to learn (i.e reach the global minima).
  + Thus a standard learner can only learn a few of the all possible functions in a hypothesis space.
* Correlation between two variables does not necessarily imply causation between them.
  + Often times there is a third variable which is causing the two variables to correlate to each other.
  + Machine learning is generally based on the observational data, where unlike the experimental data, the predictive variables are not under control of the learner.
  + Observing pattern between two set of variables, say price of house purchased and price of car bought might imply a correlation between the two variables but neither of the variable is causing the other to happen; while one might say that it could have a common cause which is the income of the individual.