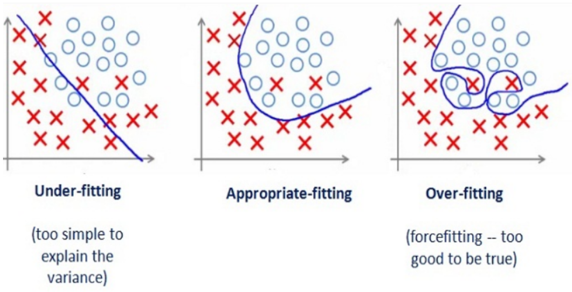
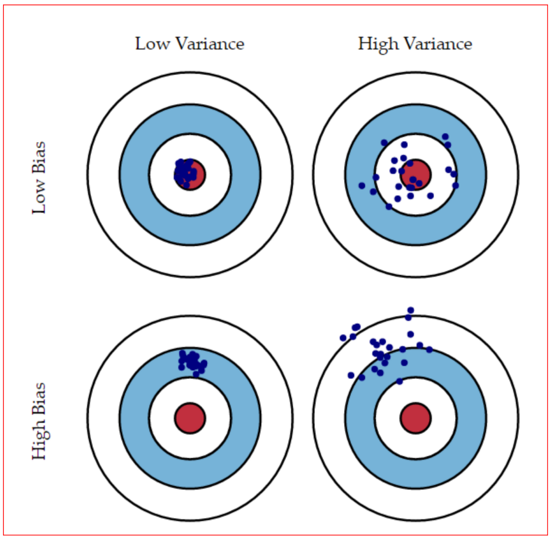
ML Notes

LECTURE 1

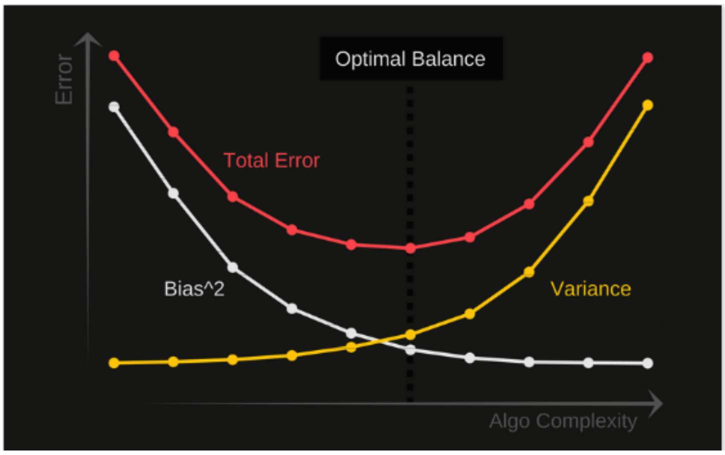
- **What are the Supervised and Unsupervised and Reinforcement Learning ?**
    - **Supervised Learning**. Supervised learning algorithms create a mathematical model for a set of data that contains both the inputs and the desired outputs. We train the machine using data which is **"labelled**." It is like the learning process which takes place in the presence of a supervisor. A supervised learning algorithm learns from labeled training data, helping you to predict outcomes from new data.
        - **Examples**: *Supervised Learning is used in* Linear regression for regression problems, Random forest for classification and regression problems, Support vector machines for classification problems.
        - **Algorithms:** K-Nearest Neighbours, SVM, Linear Regression, Logistic Regression, Decision Trees, Naive Bayes, Neural Networks.
        - **Applications**: Natural language processing, computer vision, data analytics
            - Image Recognition
            - Image Segmentation
            - Sentiment Analysis
            - Text Prediction on Search Engines and etc.

    - **Unsupervised Learning**. Unsupervised learning algorithms take a set of data that contains only inputs, and identify structure in the data, like grouping or clustering of data points. Unsupervised learning is a machine learning technique, where you need not supervise the model. Instead, you need to allow the model to work on its own to discover information. It mainly deals with the unlabelled data. Unsupervised learning algorithms allow you to perform more complex processing tasks compared to supervised learning.
        - **Applications**: A common application of Unsupervised Learning is the Recommendation System.
        - **Algorithms:** Principal component analysis (PCA), K- Means, Singular value decomposition, Apriori algorithm for association rule learning problems
    - **Reinforcement Learning.** Reinforcement learning is **a machine learning training method based on rewarding desired behaviors and/or punishing undesired ones**. In general, a reinforcement learning agent is able to perceive and interpret its environment, take actions and learn through trial and error.
- **Classification and Regression**
    - **Classification.** In classification, the model is trained in such a way that the output data is separated into different labels (or categories) according to the given input data. It can be divided by 2 - **Binary classification and Multiclass Classification.**
    - **Regression.** Unlike classification, here the regression model is trained in such a way that it predicts continuous numerical value as an output based on input variables.

- **Underfiting & Overfitting**
  - Overfitting and underfitting can be explained using below graph:
  - 

  - **Underfiting - If MSE of test and MSE of train dataset is very high.**By looking at the graph on the left side line does not cover all the points shown in the graph. It also called High Bias.
    - **Solving Underfitting problem** - To solve the problem of Underfitting we can use polynomial regression. The training error will tend to decrease as we increase the degree d of the polynomial. At the same time, the cross-validation error will also decrease to some extent
  - **Overfitting - When training Error is small and the test error is large.** Where as the graph on right side, shows the predicted line covers all the points in graph. But it is not true, coz it line covers noise and ouliers. It is also called High Variance.
    - **Solving Underfitting problem:**
      - **Adding more data**
      - **Reduce the complexity of model**
      - **L1, L2** regularizations
- **Bias and Variance**
  - Presence of bias or variance causes overfitting or underfitting of data.
  - **Bias.** Bias is how far are the predicted values from the actual values. If the **average predicted values are far off from the actual values then the bias is high**. High bias causes algorithm to miss relevant relationship between input and output variable. When a model has a high bias then it implies that the model is too simple and does not capture the complexity of data thus **underfitting the data**.
  - **Variavce**. Variance occurs when the model performs good on the trained dataset but does not do well on a dataset that it is not trained on, like a test dataset or validation dataset. **Variance tells us how scattered are the predicted value from the actual value**.
  - 

- **Bias and Variance Trade off**

- **Why Bias Variance Tradeoff?** If our model is too simple and has very few parameters then it may have high bias and low variance. On the other hand if our model has large number of parameters then it's going to have high variance and low bias. So we need to find the right/good balance without overfitting and underfitting the data. This tradeoff in complexity is why there is a tradeoff between bias and variance. An algorithm can't be more complex and less complex at the same time.
- **Total Error.** To build a good model, we need to find a good balance between bias and variance such that it minimizes the total error.
  - **Total Error = Bias^2 + Variance + Irreducible Error**
  - 

  - An optimal balance of bias and variance would never overfit or underfit the model. Therefore understanding bias and variance is critical for understanding the behavior of prediction