1. **What are differences between AI and ML?**

* ML is a subset of AI
* Data is essential part of ML
* ML is not AI; ML is not Logic or Rule Satisfaction.
* ML is not biology or Neuroscience (Neural Network or Inspiration from Biology).
* ML research are cutting edge of AI research. AI uses ML to do certain task, but it doesn’t have to be ML always.

1. **What Mathematics is needed for AI?**

* Gradient Descent (Understanding of Multivariate Calculus).
* Probability Calculations (Integrals).
* Bayesian Inference.
* Statistics and Probability Theory (Bayes Rule).
* Linear Algebra.

1. **Real world example of semi-supervised learning?**
2. **What is Hold out data set, apart from Train and test set? What is its purpose?**
3. **What are Type-1 and Type-2 errors?**

**Type I error** is the “False Positive” error. For example, when a defendant is not guilty, but was declared guilty by the jury or judges. The probability of this error is denoted by Alpha.

**Type II error** is the “False Negative” error. Building on the above example, a defendant is guilty, but was declared as innocent by the jury or judges. The probability of this error is denoted by Beta.

1. **What is in Sample Error and Out of Sample Error?**

* In Sample Error**:** The error you get on same data that you used to build predictor, sometimes called re-substitution error.
* Out of Sample Error: The error you get on new data set, sometimes called generalization error.
* In Sample error < out of sample error
* Overfitting means too much difference between in Sample error and out of sample error
* We need to care about out of sample error.
* We can always design a perfect in sample predictor but that will not perform well on new data.

1. **What is sampling (Over Sampling and Under Sampling), Explain with example?**

Given a set of dataset with 6 males and 3 female I can get a balanced sample by:

Under sampling: 3 Female + 3 out of 6 Male.

Oversampling: 6 Male + twice of 3 Female.

For over sampling us use a technique called SMOTE (Synthetic minority oversampling technique).

It creates new vectors between existing data points.

1. **What are some exmaples of Classification Problem?**

* Credit Scoring. (Non-machine learning Discriminant analysis).
* Predicting probability.
* Pattern recognition (OCR, Speech Recognition, Face Recognition, Medical Diagnosis)
* Knowledge Extraction (Learning a rule from data, properties of low risk and high risk customers).
* Compression
* Outlier Detection

1. **Unsupervised Learning:**

* We have input data and no Idea what we are looking for. Goal: Find regularities in input
* Density Estimation
* Clustering (Based on clusters, company may decide strategies to target specific group).
* Document Clustering

1. **Reinforcement** **Learning**: Output of the system is a sequence of actions. The policy, the sequence of correct actions to reach the goal is what is important.

* Game Playing: (Chess etc.)
* Robot Navigation:
* OpenAI gym