Question 1 - Data Science, Artificial Intelligence and Big Data

1. What is studied in Data Science?

A. Data Science is an interdisciplinary field that involves studies from the intersection of the fields namely, Mathematics & Statistics, Machine Learning, Computer Science, Data Processing, and Domain knowledge from the problem under consideration.

Hence, a) Predictive modelling(from the domain of machine learning involving deep learning, gradient boosting, SM, regression modeling)

b) Hypothesis testing (from the domain of Statistics)

c) Pattern Discovery (from the domain of Statistics and Machine Learning involving clustering, dimensionality reduction) are typically studied in order to gain actionable insights into the latent relationships existing among the data, for making predictions and for developing expert / recommender systems.

2. What is the difference between Data Science and AI?

A. Some of the key differences between Data Science and AI are as follows:

a) Data Science has techniques for various kinds of processing of data that involves data cleaning, exploratory analysis and visualization of the results. On the other hand, AI involves predictive modeling for future trends.
b) Data Science involves rigorous statistical techniques to be applied on data in its entire life cycle. On the other hand, AI involves algorithmic procedures to implement the predictive models mentioned above.
c) It is mainly concerned with Data Analytics whereas Ai is concerned with machine learning and deep learning techniques.
d) Data Science does not involve a high degree of scientific processing but AI involves a high degree of scientific processing.
e) Data Science is about deciphering the hidden meaning of data whereas AI develops autonomy which is used in reasoning, planning and cognition.

3. What is Big Data? Provide some examples for Big Data.

A. Big Data refers to a more larger and complex data sets from heterogeneous data sources which posses the following the following five characteristics:

a) Volume - The size of the data is huge and requires a huge reliable storage.
b) Variety - The data possesses a high and varying number of features.
c) Velocity - The data requires high real-time processing speed
d) Variability - The data is highly disparate and has multitude of dimensions owing to the multiple data sources.
e) Veracity - the data is noisy and has biases and abnormalities.

Examples of big data are biological assay data, bioimages, sensor data from electronic devices, Electronic Health records, Patient Blogs, etc.

4. What is the challenge with -omics data?

A. The challenges of omics data are as follows:

a) The data represents typically varied molecular features and thus deciphering the meaning of varied omics data in a particular context is difficult.
b) A typically omics data will have a very high number of features as compared to a low number of samples.
c) The integration of omics data from different modalities is quite challenging because of its qualitative nature.
d) Similarly, integration of omics data with nono-omics data in a relevant context is also quite challenging.
e) Omics data are inherently quite noisy and variable. Thus, it requires efficient pre-processing for further use in the analysis pipeline.