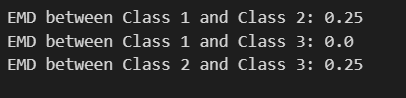
1 (e)

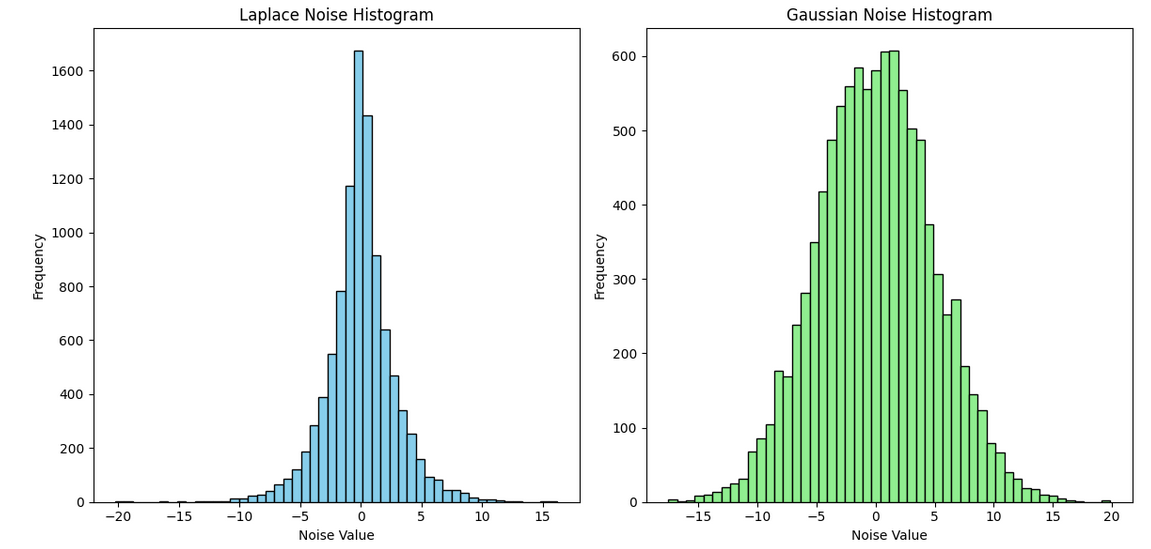
The code calculates the Earth Mover's Distance (EMD) between disease distributions in three classes (Class 1, Class 2, and Class 3) using a distance matrix that defines the "cost" of transforming one disease into another. Each class has a distribution of three diseases (Heart Disease, Viral Infection, Cancer). The emd() function computes the minimum work required to transform one distribution into another. It then prints the EMD between Class 1 and Class 2, Class 1 and Class 3, and Class 2 and Class 3. The smaller the EMD value, the more similar the distributions are.



2 (c)

The code generates and compares Laplace and Gaussian noise distributions used in differential privacy. It defines functions to generate Laplace noise based on l1​-sensitivity and Gaussian noise based on l2-sensitivity, with specified privacy parameters epsilon and delta. Noise samples of size 10,000 are drawn from each distribution. Two histograms are plotted side by side to visualize the frequency distribution of the generated noise values. The Laplace noise has a sharper peak and heavier tails compared to the Gaussian noise, reflecting their different sensitivities and scaling factors.

The Gaussian mechanism typically provides more utility compared to the Laplace mechanism when delta > 0, especially in high-dimensional data, as it adds less noise for the same privacy guarantee. In the code, the Gaussian noise histogram shows a more concentrated distribution around zero, indicating lower noise variance, which leads to better utility in practice.



2 (d)

Simple Composition: epsilon\_total=k\*epsilon

Advanced Composition: epsilon\_total=sqrt(2k\*log(1/delta))\*epsilon+k\*epsilon\*(exp(epsilon)-1)

The code calculates and plots the cumulative privacy loss (ε) for Laplace and Gaussian mechanisms under simple and advanced compositions for up to 100 queries. Simple composition accumulates privacy loss linearly with each query, while advanced composition uses a more complex formula to reduce the cumulative loss. Gaussian composition includes both epsilon and delta parameters, representing additional privacy constraints. The code generates two plots, showing how cumulative privacy loss increases with more queries for both mechanisms. Advanced composition provides lower cumulative privacy loss compared to simple composition, especially for a large number of queries.

