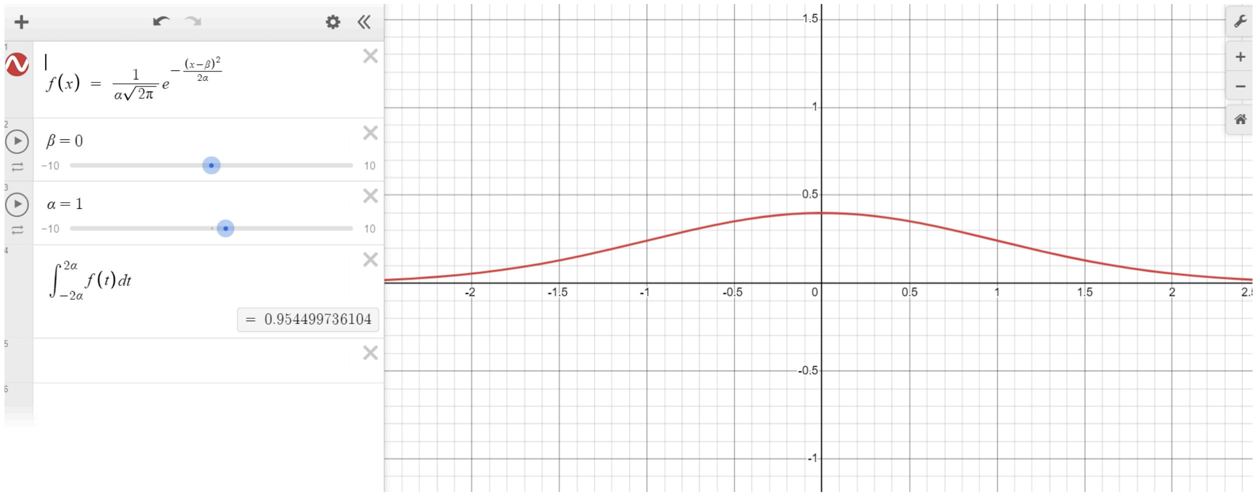
For a particular point say x, the majority of values of y obtained were in range (x-2,x+2) and this is because the random number from gaussian distribution which is added has about 95% probability to be in between -2σ to 2σ



USES OF GENERATIVE MODELS:

- Generative models can create realistic synthetic data, useful in scenarios where collecting real data is challenging, expensive, or time-consuming
- Generative models can create new training samples to augment the dataset, improving the robustness and performance of predictive models.
- By learning the normal distribution of data, generative models can identify outliers or anomalies. This is useful in fraud detection.
- Generative models can simulate various scenarios and outcomes based on certain conditions, aiding in decision-making processes, risk assessment, and policy planning.

REPORT ON MLE

Maximum Likelihood Estimation(MLE) is an algorithm used to estimate the parameters, given the data such that by using those parameters the given data is most probable. This algorithm assumes that the data is composed of Independent and Identically Distributed(I.I.D) samples.

We denote $f(X/\theta)$ to refer to PMF or PDF of given data. Hence, the likelihood of data is given as :

$$L(\theta) = \prod_{i=1}^{n} f(X_i/\theta)$$

In order to maximize $L(\theta)$ we use MLE and represent the best choice of parameters as θ' i.e.

$$\theta' = argmax\, L(\theta)$$

Since logarithm is a monotonic function we apply MLE on log of $L(\theta)$ as log makes calculation easier. Following are some examples to demonstrate its implementation:

Normal I.I.D.

Say we are given n samples $X_1, X_2, \ldots, X_n$ such that for all i, $X_i \sim N(\mu, \sigma^2)$, hence the parameters here to estimate are μ and σ^2.

$$L(\theta) = \prod_{i=1}^{n} f(X_i/\theta)$$

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}}$$

$$log\, L(\theta) = \sum_{i=1}^{n} [- \, log\, (\sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}(X_i - \mu)^2$$

Now we calculate the first derivative of above eq^n w.r.t μ and σ^2 and equate them to 0 to get the values of μ and σ^2 i.e.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

Linear Transform with Noise

Similar to previous case say we are given n samples $(X_1, Y_1), (X_2, Y_2), \ldots\ldots, (X_n, Y_n)$ and the model be $Y = \theta X + Z$ where Z is the noise and $Z \sim N(0, \sigma^2)$.

Also $\theta X + Z$ is sum of gaussian and a number hence one can say that $Y/X \sim N(\theta X, \sigma^2)$ and we need to find θ to maximize the probability of occurrence of given data. Hence,

$$f(Y_i/X_i, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y_i-\theta X_i)^2}{2\sigma^2}}$$

$$L(\theta) = \prod_{i=1}^{n} f(Y_i, X_i/\theta) = \prod_{i=1}^{n} f(Y_i/X_i, \theta) f(X_i)$$

$$log\, L(\theta) = -\, nlog(\sqrt{2\pi}) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \theta X_i)^2 + \sum_{i=1}^{n} f(X_i)$$

Since $f(X_i)$ is independent of θ to maximize $log\, L(\theta)$ we need to maximize $-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \theta X_i)^2$

$$\theta' = argmax - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \theta X_i)^2$$

$$\theta' = argmin \frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - \theta X_i)^2$$

This shows that for such a data the best possible value of θ is one which minimizes the squared error of prediction of Y.

Report on Generative Model

The purpose of our generative model is to remove the bias field in the MRI images and cluster the tissues in image based on their intensities observed. Since we need to identify gray matter, white matter and cerebrospinal fluid so we need to cluster voxels in 3 clusters(hence k=3). But for proper clustering we need to remove the bias field as it alters the intensities observed from actual intensity.We are able to do so as the bias field is spatial smooth. At a particular voxel i:

$$y_i = x_i b_i + n_i \text{ where,}$$

y_i : Observed Intensity , x_i: Actual Intensity , b_i: Bias Field , n_i: Noise and $n_i \sim G(0, \sigma^2)$

We represent c_k as the uncorrupted intensity of each tissue.

For a voxel i and a neighborhood voxel j, if they both belong to same tissue/ cluster then,

$$x_j = c_k$$

$$b_i \approx b_j$$

$$\Rightarrow x_j b_j \approx c_k b_i$$