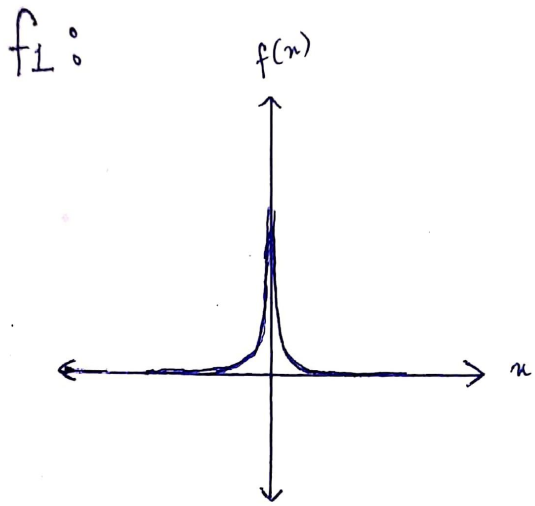
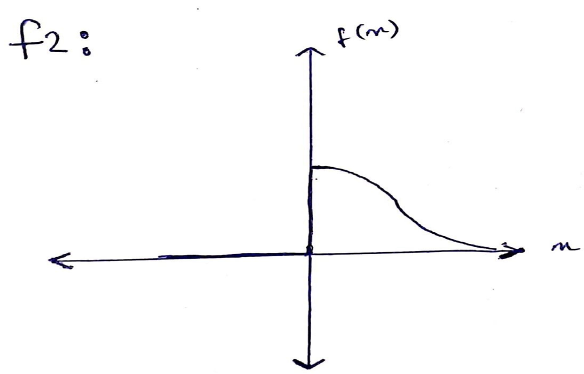
Learning Rate

Given two probability density functions f_1 and f_2 described below:

f1: [Delta Function]



f2: generates only positive real numbers.



$$\int_{-\infty}^{\infty} x f_2(x) dx = \frac{1}{n}$$

$$\int_{-\infty}^{\infty} (x - \frac{1}{n})^2 f_2(x) dx = \frac{1}{n^2}$$

Wherever used: n = 1e5, d = 1e4.

A Dataset $X \in R^{n \times d}$ is generated in the following way

For i in range(n):

For j in range(d):

Choose pdf p randomly from f1 and f2. Draw a sample from p and assign it to X[i,j].

$$y \in R^n = [1, 1, 1, \ldots, 1, 1, 1]^T$$

Shyam uses linear regression to find out $W \in R^d$ such that $XW \approx y$. Note that Shyam doesn't do any feature processing on X.

Shyam's Code:

Initialize W to some random value(call it W_0).

While Converges:

$$W = W - \alpha X^T (XW - y)$$

Your task is to find out approx α such that training is as fast as possible for any given $W_0 \neq W_*$, where $W_* = (X^T X)^{-1} X^T y$.

Evaluation:

1. Participants should submit their detailed approach in pdf format. No marks will be rewarded without the approach. You can send the scanned copy also.
2. This is a theoretical question and you have to describe approach and theorems which are used in the solution.