Fourth Question: Suppose that X is Gaussian with 0 mean and variance σ_X^2 as before, but now Y is discrete with pmf

$$p_Y(y) = \frac{1}{2}; y = \pm\sigma_Y.$$

Find an upper bound for $\overline{d}(F_X, F_Y)$ that is strictly better than the bound of the First Question.

Useful fact: X Gaussian, 0 mean, variance σ_X^2

$$E[|X|] = E[X \mid X \geq 0] = \sqrt{\frac{2}{\pi}}\sigma_X$$

Solution: To get an upper bound, a joint distribution is needed on X, Y with the correct marginals. The first question used a simple product distribution, making the random variables independent. If we want a close mean squared error match, however, we want the two random variables to be correlated, in fact we want them to be maximally correlated if we want to try to achieve the lower bound of Question 2. There are several ways to create a joint distribution with the desired properties. One way is to define Y as a deterministic function of X with

$$Y = \begin{cases} \sigma_Y & \text{if } X \geq 0 \\ -\sigma_Y & \text{otherwise} \end{cases}$$

the resulting joint distribution yields the desired marginals and using conditional expectation results in

$$\begin{aligned} E[(X-Y)^2] &= \sigma_X^2 + \sigma_Y^2 - 2E(XY) \\ &= \sigma_X^2 + \sigma_Y^2 - 2\left[E(XY \mid X \geq 0)\frac{1}{2} + E(XY \mid X < 0)\frac{1}{2}\right] \\ &= \sigma_X^2 + \sigma_Y^2 - 2\sqrt{\frac{2}{\pi}}\sigma_X\sigma_Y \end{aligned}$$

This is not as good as the lower bound of Question 2, but nonetheless it turns out to actually solve the minimization. Here the joint distribution results from one random variable being a deterministic function of the other, and the operation used here is simply a binary quantizer.
Another way to get the same joint distribution is to use a classic result often derived in elementary probability classes. If U is a uniform distribution on $[0, 1]$, then the random variable $F_X^{-1}(U)$ will have cdc F_X, where F_X^{-1} denotes the inverse cdf. Thus given a single uniform U, one can generate random variables with the correct marginals via $(X, Y) = F_X^{-1}(U), F_Y^{-1}(U)$. This results in the same joint distribution as using the quantizer, and hence yields the same bound. But everyone who made a guess chose the previous approach, which amounts to a binary quantizer.