Last Question: I did not expect anyone to get this far, but two people did.
Does the lower bound you found in the previous part actually solve the minimization? That is, does the lower bound equal the distance?

Solution: You might think the bound does not yield the maximum correlation and hence the minimum "distance" since it does not achieve the bound given by Cauchy-Schwartz in Question 2, but it turns out that it is the minimum and the lower bound of Question 2 is not achievable in this nonGaussian example. Intuitively, you can not make X and Y with the given distributions any more correlated then matching their signs (in this example).
To prove that our new bound actually yields the distance, we need to show that for *any* joint distribution *with the given marginals*, it must be true that

$$E(XY) \leq \sqrt{\frac{2}{\pi}}\sigma_X\sigma_Y,$$

the value we actually achieved by a specific joint distribution. While Cauchy-Schwartz is still true here, it is too optimistic, it is not achievable. We need a better bound to the maximum correlation. This question was intended to elicit thoughts on how such an inequality might be proved for this case. One way is to use the method of indicators.
Define the indicator function

$$1(X \geq 0) = \begin{cases} 1 & X \geq 0 \\ 0 & X < 0 \end{cases}$$

and define the other indicator $1(X < 0)$ similarly. Since $1 = 1(X \geq 0) + 1(X < 0)$, we have that

$$\begin{aligned} E(XY) &= E[XY(1(X \geq 0) + 1(X < 0))(1(Y \geq 0) + 1(Y < 0))] \\ &= E[XY1(X \geq 0)1(Y \geq 0)] + E[XY1(X \geq 0)1(Y < 0)] \\ &\quad + E[XY1(X < 0)1(Y \geq 0)] + E[XY1(X < 0)1(Y < 0)] \\ &\leq E[XY1(X \geq 0)1(Y \geq 0)] + E[XY1(X < 0)1(Y < 0)] \end{aligned}$$

since the removed terms are negative. Because Y is binary, the right hand side is

$$\sigma_Y E[X1(X \geq 0)1(Y \geq 0)] - \sigma_Y E[X1(X < 0)1(Y < 0)] = \\ \sigma_Y E[X1(X \geq 0)1(Y \geq 0)] + \sigma_Y E[-X1(X < 0)1(Y < 0)].$$

Again the terms in the brackets are nonnegative and indicator functions are bound above by 1, so we have

$$E(XY) \leq \sigma_Y E[X1(X \geq 0)] + \sigma_Y E[-X1(X < 0)] = \sigma_Y E(|X|) = \sqrt{\frac{2}{\pi}}\sigma_X\sigma_Y$$

as needed. This proves the bound is actually achieved and hence yields the transportation distance. I did not expect anyone to actually go through this (and there is probably a shorter proof), I was only looking for ideas on how to decompose the expectation using the structure of the given distributions.