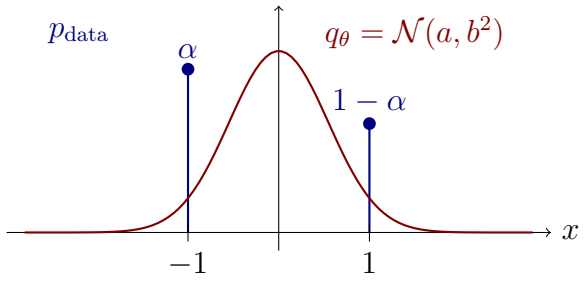
Name: _____ Total: 15 points

Let $X \sim p_{\text{data}}$ be a random variable that takes values on $\{-1, 1\}$ with $p_{\text{data}}(\{-1\}) = \alpha$ and $p_{\text{data}}(\{1\}) = 1 - \alpha$. Consider a class of Gaussian model densities $q_\theta = \mathcal{N}(a, b^2)$, where the set of parameters $\theta = (a, b)$ corresponds to the mean and standard deviation of the Gaussian.



1. Consider maximizing the likelihood of i.i.d. samples $\{x_i\}_{i \in [m]}$ from p_{data}. That is, let

$$(a^*, b^*) = \theta^* := \arg\max_\theta \frac{1}{m} \sum_{i=1}^{m} \log q_\theta(x_i).$$

   As $m \to \infty$ and $\alpha = 0$, q_{θ^*} has zero variance. True or False? Justify your answer. (2 points)

2. When $m \to \infty$, for no value of $\alpha \in [0, 1]$ does q_{θ^*} match p_{data} exactly. True or False? Justify your answer. (2 points)

3. Let the source distribution be p_{data} and the target be q_{θ^*}. Does an optimal transport map (solution of the Monge problem) exist with the squared Euclidean distance as the cost function, for any $\alpha \in [0, 1]$? (1 point)

4. The figure below shows the optimal transport map from q_{θ^*} to p_{data} for two different values of α in the left and right plots. Which is greater, α_1 or α_2, and why? (2 points)

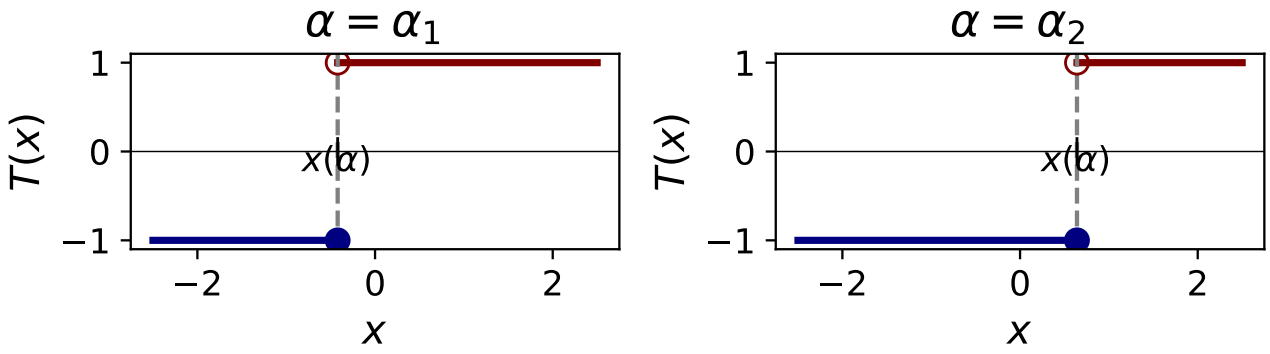


Figure 1: Optimal transport map from q_{θ^*} to p_{data}.

5. In Figure 1, the point of discontinuity in the OT map is marked $x(\alpha)$. What is the relationship between $x(\alpha)$ and α? Use the notation $\Phi_{a,b}(x)$ for the CDF of a Gaussian with mean a and variance b^2. (2 points)

6. The figure below shows the transport plans (solution of the Kantorovich problem) for the source q_{θ^*} and target p_{data} with different values of α. Blue indicates regions of low joint density and red indicates high density. The four values of α used are $\{0.01, 0.2, 0.7, 0.9\}$, but these are not in order! Unscramble them (2 points for correct order, 2 points for explanation):

$$\alpha_1 = \ldots\ldots, \quad \alpha_2 = \ldots\ldots, \quad \alpha_3 = \ldots\ldots, \quad \alpha_4 = \ldots\ldots$$

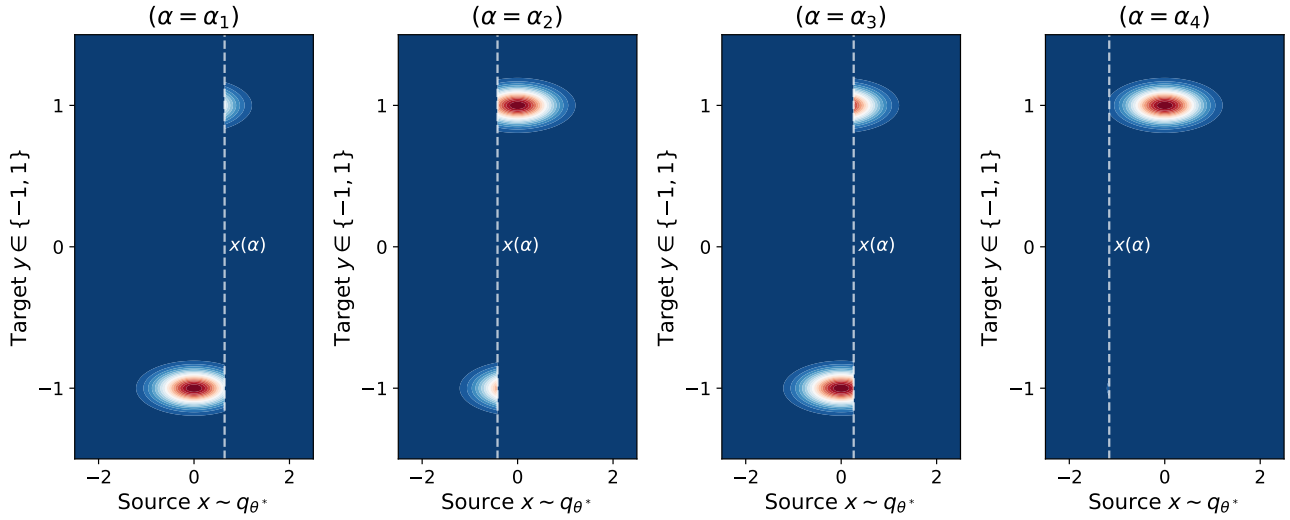


Figure 2: Transport plans from q_{θ^*} to p_{data}.

7. Using the above plots or otherwise, explain for what value of α would q_{θ^*} be the worst generative model for p_{data} in terms of the W_2 (Wasserstein-2) metric. (2 points)