MDL - TD6

27/02/2024

Exercise 1: Distances between distributions

Q1: Show that the total variation distance $d_{\mathrm{TV}}(\mu,\nu)$ defined as $d_{\mathrm{TV}}(\mu,\nu) = \|P - Q\|_1$ where P and Q are the densities of μ and ν, also verifies:

$$d_{\mathrm{TV}}(\mu,\nu) = \max_{f \in L^\infty} \mathbb{E}[f(X) - f(Y)]\,,$$

where $(X,Y) \sim (\mu,\nu)$ and L^∞ is the set of functions bounded by 1.

Q2: Show that the Wasserstein distance $d_{\mathrm{W}}(\mu,\nu)$ defined as $d_{\mathrm{W}}(\mu,\nu) = \max_{f \in L^1} \mathbb{E}[f(X) - f(Y)]$ where L^1 is the set of 1-Lipshitz functions, verifies:

$$d_{\mathrm{W}}(\mu,\nu) \leq \inf_{(X,Y)\in\Pi(\mu,\nu)} \mathbb{E}(\|X - Y\|)\,,$$

where $\pi(\mu,\nu)$ is the set of couplings (X,Y) such that $X \sim \mu$ and $Y \sim \nu$.

Q3: Let μ be the law of the uniform random variable over the hypercube $\{0,1\}^d$. For $X_1,\dots,X_n \sim \mu$, let $\hat{\mu}_n$ be the empirical distributions over these n samples. Show that

$$d_{\mathrm{W}}(\mu,\hat{\mu}_n) \geq 1 - \frac{n}{2^d}\,.$$