APMTH 221 FINAL PROJECT PROPOSAL

JIAFENG (KEVIN) CHEN AND FRANCISCO RIVERA

1. Collaboration

Francisco and Kevin will work together to research the relevant optimization techniques helpful to calculating the Wasserstein distance quickly. Then, they will split up techniques, and through implementations of them, empirically (and when possible, theoretically) assess their performance. Time permitting, they will then attempt to create a new algorithm that performs competitively with state of the art benchmarks.

2. Problem Statement

If we are given two probability measures μ and ν on metric space (M, d) with finite moments,

$$\int_M d(x, x_0)^p d\mu(x) < \infty \text{ and } \int_M d(x, x_0)^p d\nu(x) < \infty$$

then we define the pth Wasserstein distance as

$$W_p(\mu, \nu) := \left(\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{M \times M} d(x, y)^p \mathrm{d}\gamma(x, y) \right)^{1/p}$$

where $\Gamma(\mu, \nu)$ is the collection of all measures on $M \times M$ with marginals μ and ν.

This can also be thought of as the optimal transport problem. If we think of a probability distribution as a collection of "mass" at different points in M (with density of the mass proportional to the probability measure), then the Wasserstein distance captures the cost of transforming one distribution into the other in units of mass times distance. This analogy is particularly interpretable if we have a discrete metric space, where probability measures can be viewed as point masses.

Solving the optimization problem underpinning this metric is in general hard and remains an unsolved problem. We will aim to test the effectiveness of different optimization techniques to this problem, with particular focus on discrete metric spaces.

3. Motivation

Numerous important applications exist for the Wasserstein metric.

4. Data

Because wish to test the effectiveness of Wasserstein distance calculations for arbitrary distributions, we will generate test distributions using simulation and do not need to concern ourselves with finding real data. When necessary, we will use standard distributions found in the literature for consistency.

5. Deliverables

We will implement optimization techniques to calculate Wasserstein distances and attempt a novel algorithm that performs competitively.

Date: February 21, 2018.

6. Next Steps

- We will use linear programs to calculate Wasserstein distances for discrete metric spaces. We will explore average complexity for randomly generated test cases.
-