

Clustering Distributions with Optimal Transport Experiments

Kevin O'Connor

3/26/2019

Introduction

This document includes the code for running experiments in “Clustering Distributions with Optimal Transport” by Kevin O’Connor.

Clustering Gaussians

```
# Generate set of gaussian distributions
n <- 10 # Number of distributions
m <- 100 # Number of points in each distribution
mu.mat <- matrix(runif(n * 2, min = -10, max = 10), nrow = 2, ncol = n)
dat <- data.frame("class" = rep(1:n, each = m)) %>%
  mutate("dim1" = rnorm(m * n, mu.mat[1, rep(1:n, each = m)])) %>%
  mutate("dim2" = rnorm(m * n, mu.mat[2, rep(1:n, each = m)]))
# Plot data
ggplot(dat, aes(x=dim1, y=dim2, color=as.factor(dat$class))) +
  geom_point(show.legend = FALSE)
```

