Generative Data Obfuscation!

Peyton, Krahulik Computer Science and Mathematics
Kyle Caudle Mathematics
Randy Hoover Computer Science

Real data often contains proprietary information, or in the case of military applications, the data may actually, be classified. The goal of this research is to investigate generative models and maps that will create a distribution that resembles the population with sensitive information that has been masked. In this research, a generative adversarial network (GAN) is used to generate data from the masked distribution that is similar to the population distribution while obscuring sensitive information. Through computer experiments, we show that if we have a large sample from the population and a relatively small sample of data with masked information, we can create fake instances from the masked distribution. We measure success by comparing the distance between the masked distribution and the population distribution.