GAN for discrete dataset

1. Introduction

This paper aims to develop a GAN model which can generate synthetic data for any given real life tables.



Once the GAN model is trained, it should be able to generate any number of fake data with the same data features and distributions as the original databse table.

The purpose is to design such a GAN Model.

1. Review

Quite a few papers talking about GAN, can be found here

<https://github.com/zhangqianhui/AdversarialNetsPapers>

1. Methods Overview

A sample input looks like the table below:

Table 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Id | gender | state | salary | age |
| 1 | m | nsw | 60123 | 23 |
| 2 | f | vic | 137598 | 35 |
| 3 | m | tas | 181489 | 41 |
| 4 | m | act | 50368 | 21 |
| 5 | f | wa | 0 | 16 |
| 6 | … |  |  |  |
| … |  |  |  |  |

The current whole workflow like:

