



Generative Adversarial Network aplicada a geração de dados Sintéticos para saúde.

Lucas Mendes Barboza



Motivação

- O acesso a bancos de dados médicos limita o avanço de pesquisas e desenvolvimento de produto.
- Existem várias questões acerca da segurança dos dados e da privacidade dos pacientes.
- Dados sintéticos já vem sendo amplamente para outros propósitos.



Objetivo

Aplicar técnicas de aprendizagem de máquina para gerar dados de pacientes sintéticos



Generative Adversarial Networks (GAN)

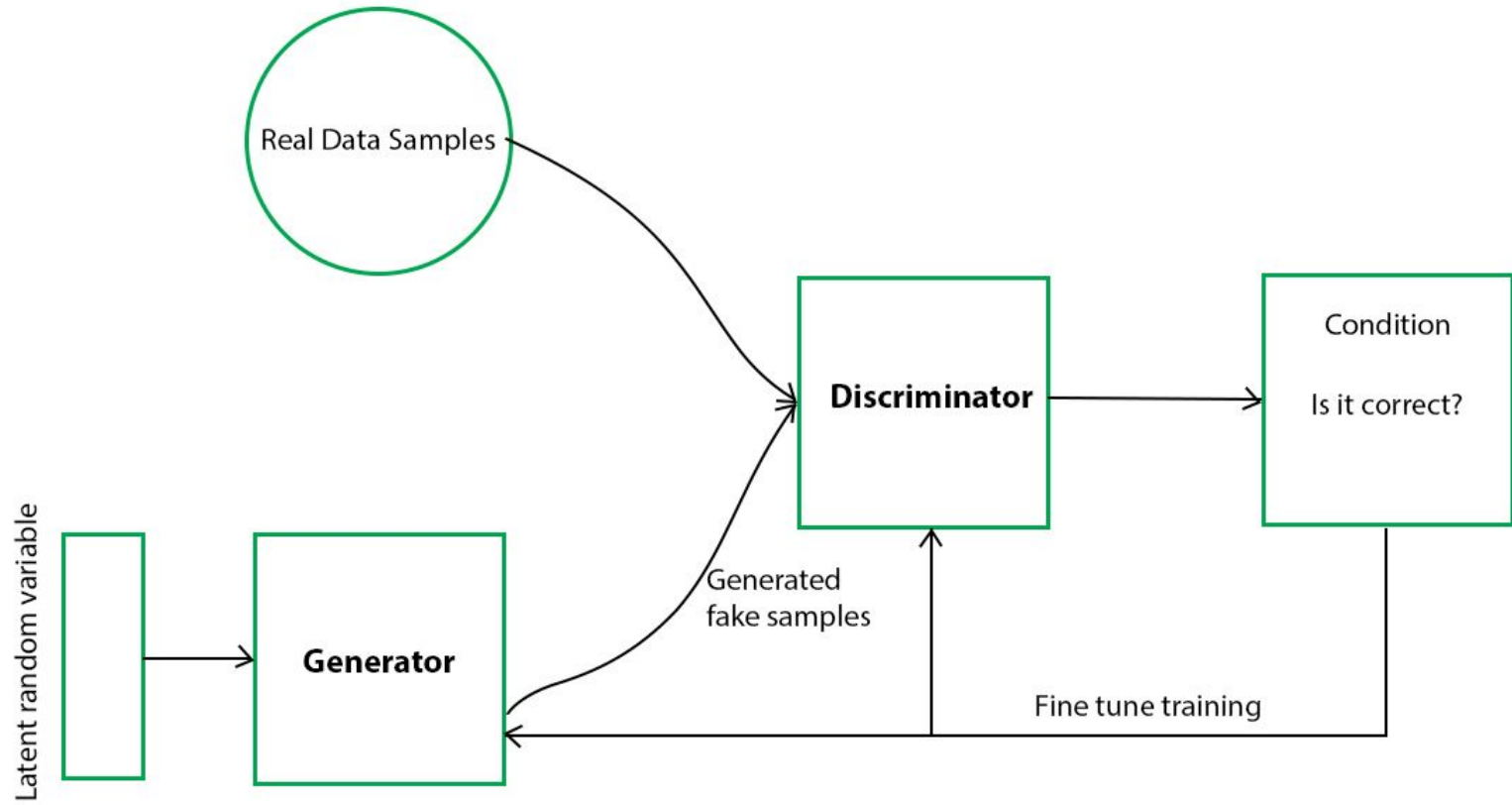
As Generative modeling (GM) é um tipo de modelo de aprendizado não supervisionado no qual o modelo automaticamente aprende padrões do dado entrada.

Já as Generative Adversarial Networks (GAN) são uma abordagem das GM's. Diferente das outras abordagens a GAN's usam aprendizado supervisionado para gerar o modelo Generativo.



GAN - Por que Adversarial?

Nas GAN's dois submodelos são treinados ao mesmo tempo, um tem o objetivo de gerar dados que sejam indistinguíveis dos reais e outro com objetivo de diferenciar dado real do falso.



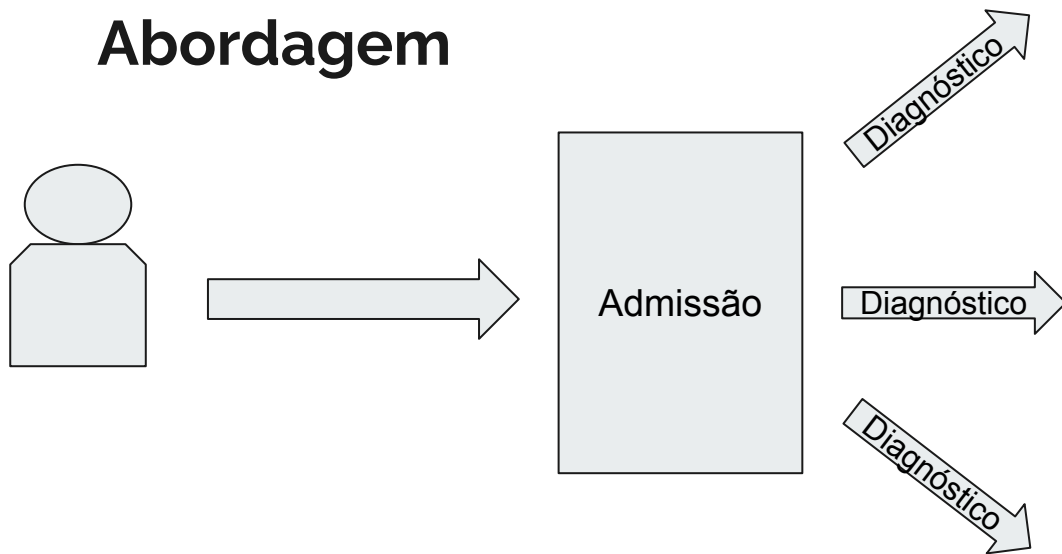


MIMIC-III

O dataset MIMIC III é o resultado de um esforço de coleta de 11 anos (2001-2012). O dataset possui mais 40 mil entradas de pacientes. Ele é composto por dados Admissão, diagnósticos e procedimentos realizados com paciente do Beth Israel Deaconess Medical Center.

Foram usados os dados de admissão com 100 pacientes únicos e os dados de diagnósticos com 581 diagnósticos únicos. Os diagnósticos no dataset estão no padrão ICD9.

Abordagem



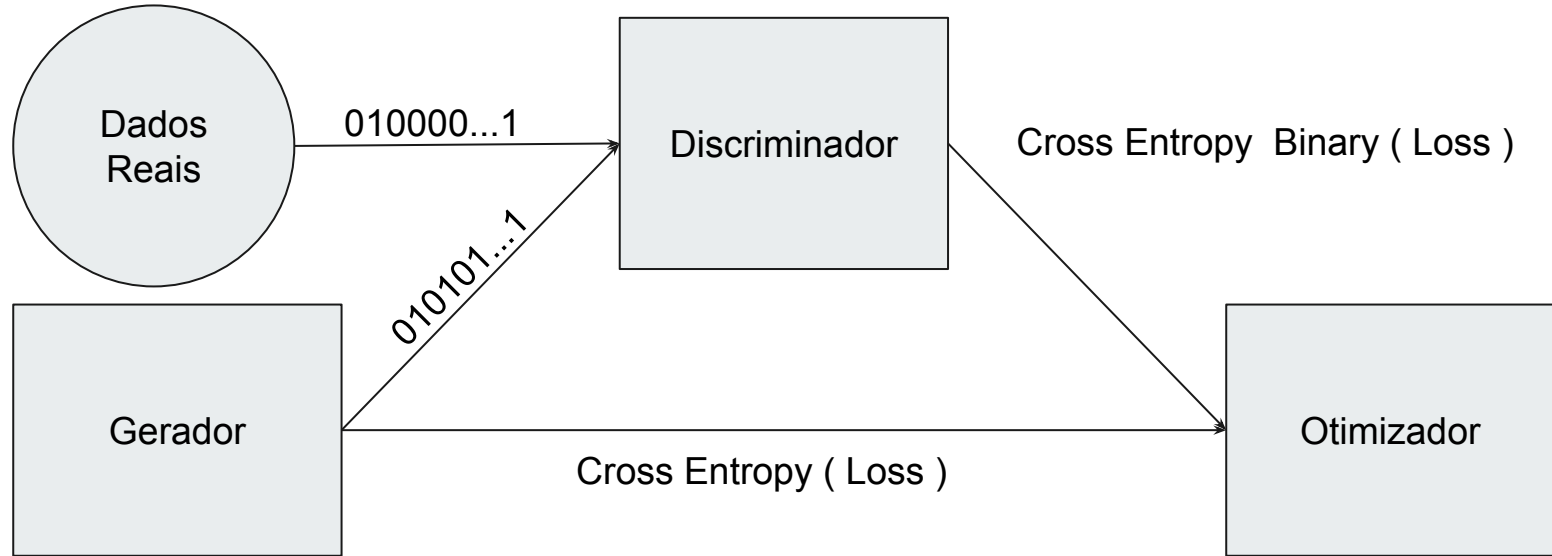
Paciente	Diagnóstico
n1	0083, E045
nn	0083, ...




Abordagem

	Diagnóstico 1	...	Diagnóstico n
1	1	...	1
2	0	...	1

GAN





```
[16] def discriminator(hsize=[10, 10],reuse=False):
    # Cria o modelo / Funciona como stack de opercoes
    model = tf.keras.Sequential()
    # Cria primeira camada da rede neural usando os input Z / layer.dense cria uma rede neural do tipo densa
    model.add(tf.keras.layers.Dense(hsize[0],name='DL1' ,activation='sigmoid')) #,input_shape=(BATCH_SIZE,581)))
    # Cria a segunda camada da rede usando o input h1(Z)
    model.add( tf.keras.layers.Dense(hsize[1], name='DL2',activation=tf.keras.layers.LeakyReLU()))
    model.add( tf.keras.layers.Dense(hsize[1],name='DL3', activation=tf.keras.layers.LeakyReLU()))
    # Cria a de saída camada da rede usando o input h2(h1(Z))
    model.add(tf.keras.layers.Dense(2, name='DL4'))
    # Cria a de saída camada da rede usando o input h3(h2(h1(Z)))
    model.add(tf.keras.layers.Dense(1,name='DLout', activation='relu'))
    return model
```

```
[15] def generator(hsize=[10, 10],reuse=False):  
    # Cria o modelo / Funciona como stack de opercoes  
    model = tf.keras.Sequential()  
  
    # Cria primeira camada da rede neural usando os input Z / layer.dense cria uma rede neural do tipo densa  
    model.add(tf.keras.layers.Dense(hsize[1],name='GL1' ,activation=tf.keras.layers.LeakyReLU(),)) #input_shape=(BATCH_SIZE,581)))  
  
    # Cria a segunda camada da rede usando o input h1(Z)  
    model.add( tf.keras.layers.Dense(hsize[1],name='GL2',activation=tf.keras.layers.LeakyReLU()))  
  
    # Cria a de saida camada da rede usando o input h2(h1(Z))  
    model.add(tf.keras.layers.Dense(581,name='GLout', activation='relu'))  
    return model
```





Bibliografía

<https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/>

Walonoski J, Kramer M, Nichols J, Quina A, Moesel C, Hall D, Duffett C, Dube K, Gallagher T, McLachlan S. Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. J Am Med Inform Assoc. 2018 Mar 1;25(3):230-238. doi: 10.1093/jamia/ocx079. Erratum in: J Am Med Inform Assoc. 2018 Jul 1;25(7):921. PMID: 29025144; PMCID: PMC7651916.

Baowaly MK, Lin CC, Liu CL, Chen KT. Synthesizing electronic health records using improved generative adversarial networks. J Am Med Inform Assoc. 2019 Mar 1;26(3):228-241. doi: 10.1093/jamia/ocy142. PMID: 30535151; PMCID: PMC7647178.

Dahmen J, Cook D. SynSys: A Synthetic Data Generation System for Healthcare Applications. Sensors (Basel). 2019 Mar 8;19(5):1181. doi: 10.3390/s19051181. PMID: 30857130; PMCID: PMC6427177.