

Geração de genes candidatos com as Generative Adversarial Networks (GAN's)

Resumo:

Com os dados do Ecoli, a gente faz um processado para que eles fiquem em um intervalo de [-1 até 1], isso para que a rede faça bem seus operações e possa convergir, depois, a GAN é treinada com os dados reais e os dados gerados pelo gerador (dados fake). A gente chega a um ponto de conversão com uma perda de 0.69, tendo em conta que a perda desejada é 0.5.

Metodología:

A figura 1 mostra a metodologia empregada para poder gerar os genes candidatos. Para poder mostrar as distribuições dos dados, a gente fez um PCA (*Principal Component Analysis*) de 2 componentes para poder mostrar eles em um gráfico. A figura 1 mostra a os candidatos *fake* os quais têm uma distribuição semelhante a dristribuição real.

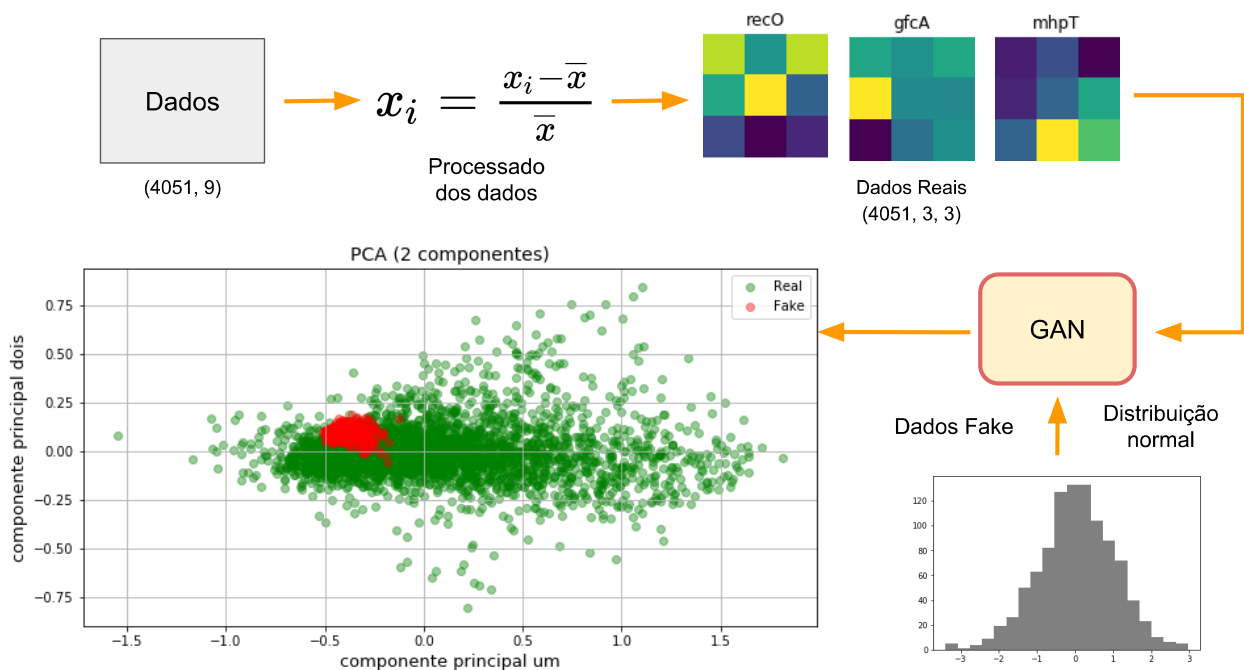


Figura 1. Metodología empregada.

A figura 2 mostra a arquitetura da rede geradora e também a arquitetura da rede discriminadora, se pode ver que a rede Geradora tem quatro camadas, as três primeiras com a função retificadora ReLu, a primeira camada tem 16 neurônios, a segunda camada tem 32 neurônios e assim até chegar a gerar assim uma saída de (*None*, 3, 3, 1), onde *None* é o número de exemplos gerados. Por outro lado, a rede Discriminadora, tem quatro camadas também, com as 2 primeiras camadas com a função retificadora ReLu, gerando assim uma saída com um só valor, se é real ou não.

Rede geradora

Layer (type)	Output Shape	Param #
dense_9 (Dense)	(None, 16)	1616
leaky_re_lu_6 (LeakyReLU)	(None, 16)	0
dense_10 (Dense)	(None, 32)	544
leaky_re_lu_7 (LeakyReLU)	(None, 32)	0
dense_11 (Dense)	(None, 64)	2112
leaky_re_lu_8 (LeakyReLU)	(None, 64)	0
dense_12 (Dense)	(None, 9)	585
reshape_2 (Reshape)	(None, 3, 3, 1)	0
Total params: 4,857		
Trainable params: 4,857		
Non-trainable params: 0		

Rede discriminadora

Layer (type)	Output Shape	Param #
dense_13 (Dense)	(None, 3, 3, 64)	128
leaky_re_lu_9 (LeakyReLU)	(None, 3, 3, 64)	0
dense_14 (Dense)	(None, 3, 3, 32)	2080
leaky_re_lu_10 (LeakyReLU)	(None, 3, 3, 32)	0
dense_15 (Dense)	(None, 3, 3, 16)	528
flatten_2 (Flatten)	(None, 144)	0
dense_16 (Dense)	(None, 1)	145
Total params: 5,762		
Trainable params: 2,881		
Non-trainable params: 2,881		

Figura 2. Arquitetura da rede.

Resultados:

A rede foi treinada com 300 *epochs* e com um *batch size* de 16 exemplos, ela levou 3.52 minutos de treinamento, a figura 3 mostra a perda da GAN, tanto do Gerador como do Discriminador por *epochs*, mostrando que a rede converge com uma perda de 0.69.



Figura 3. Perda da rede GAN.

Por último, são gerados genes *fake*, e através do PCA com 2 componentes é gerado o gráfico de distribuição para poder ver se a distribuição gerada pelo gerador é semelhante a os dados reais. A figura 4 mostra essa comparação.

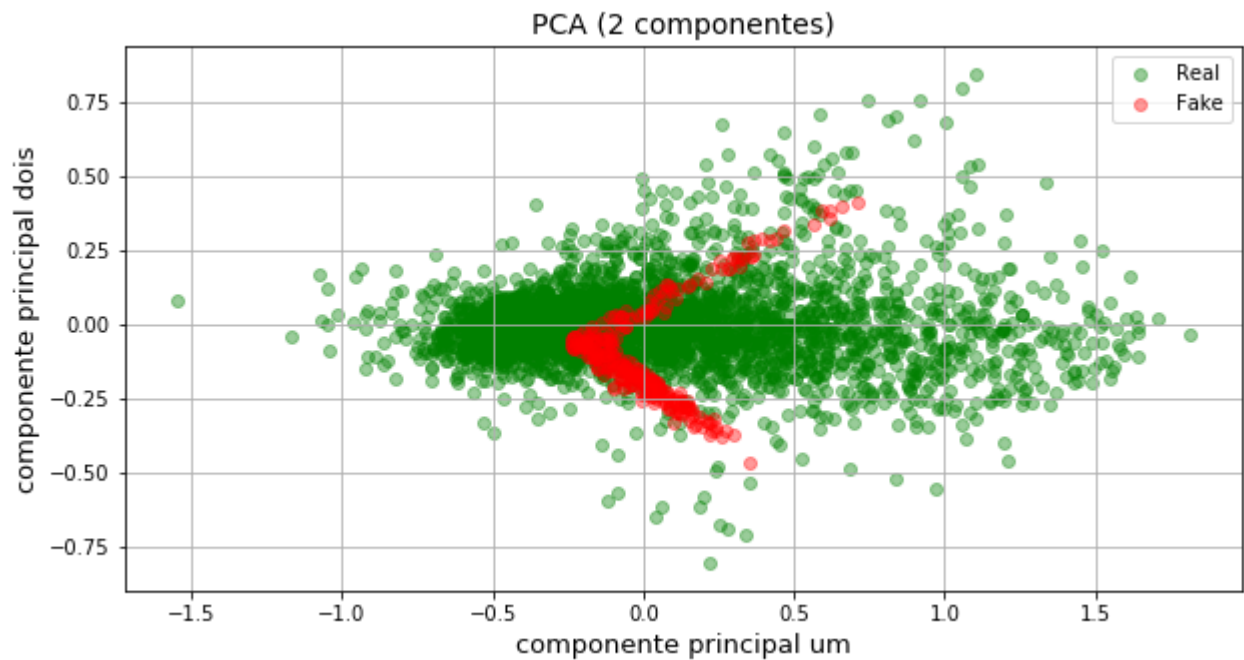


Figura 4. Comparação nas distribuições.