

Card26_Modelos_Generativos

Davi Bezerra Barros

Variational Auto-Encoders

Auto Encoders: Geralmente utilizados para compressão de imagens, remoção de ruídos e geração de dados, os autoencoders são compostos por duas partes; o Encoder, que recebe uma imagem como input e a mapeia para um espaço dimensional inferior chamado de espaço latente, e o decoder, que reconstrói a imagem original a partir desta representação espacial. Eles podem aprender a capturar as características de uma imagem enquanto descartam o ruído. Os autoencoders podem ser treinados sem supervisão e representam dados com alta dimensionalidade, e assim reduzem o espaço de armazenamento utilizado pelo dataset ao comprimir as imagens;

Variational Auto Encoders: Ao contrário dos auto-encoders clássicos, que mapeiam um dado de entrada para um ponto no espaço latente, os auto-encoders variacionais mapeiam os dados para uma distribuição probabilística em dois vetores no espaço latente; um de médias e um de desvio padrão.

General Adversarial Networks:

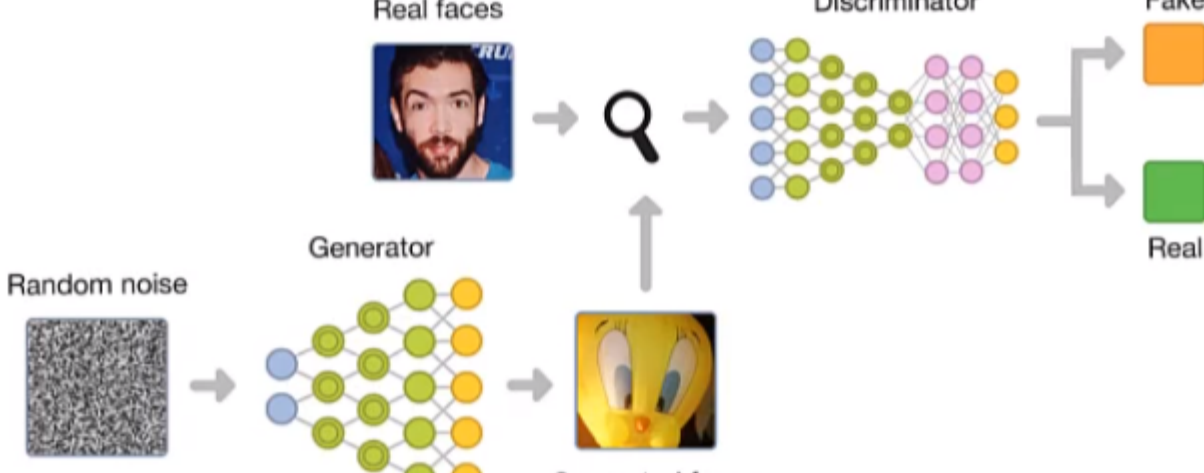
São redes neurais generativas utilizadas para criar datasets sintéticos que se assemelham a dados reais, para diferentes aplicações.

Funcionamento:

As GAN's são compostas por dois componentes principais:

Gerador: Componente responsável por criar dados falsos a partir de um vetor de ruído. Seu objetivo é criar imagens que são indistinguíveis das imagens reais.

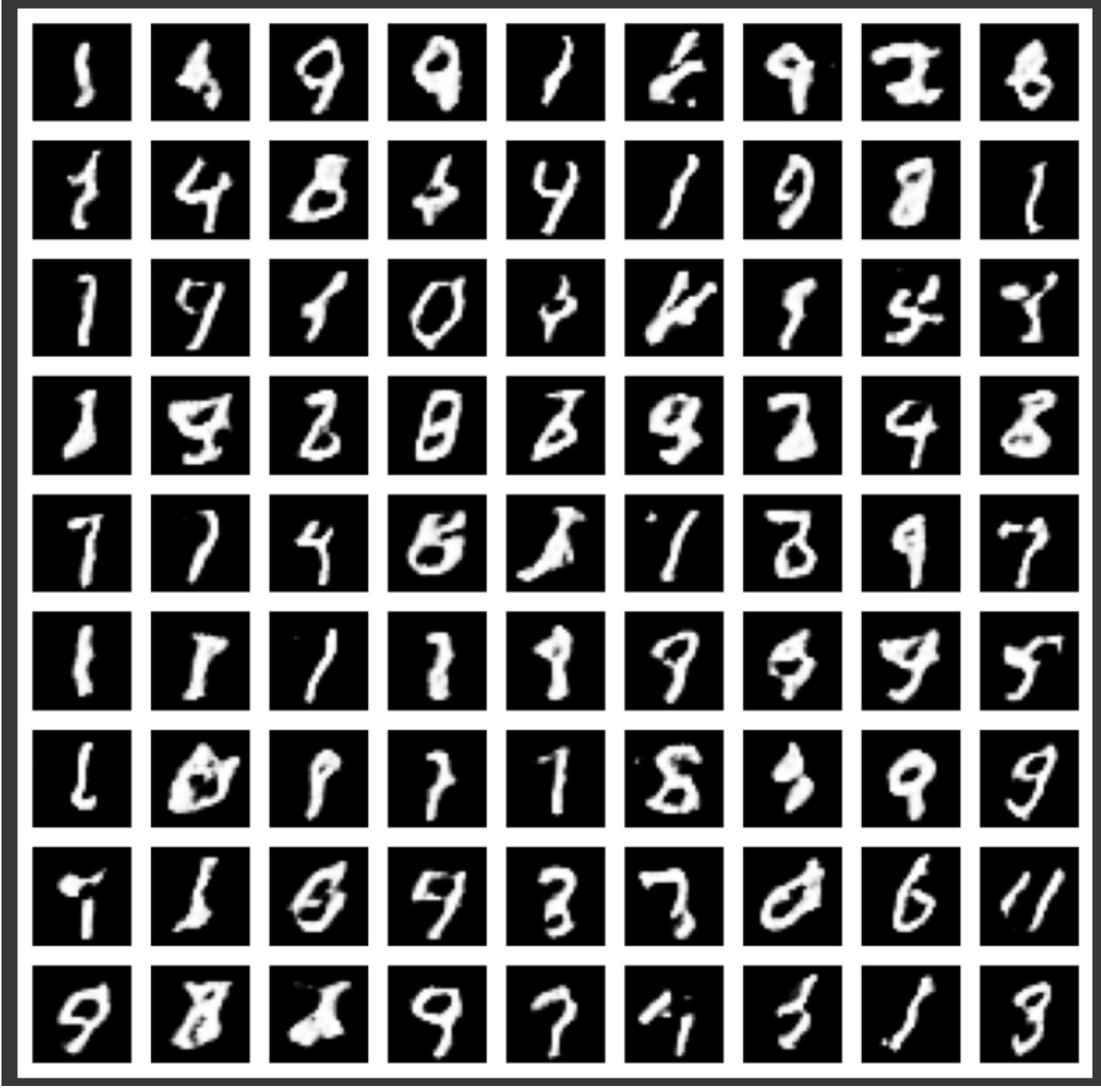
Discriminador: Recebe os dados reais e os dados falsos gerados pelo gerador, e tenta distinguir qual é o falso. O objetivo do discriminador é identificar quais imagens são falsas, e melhorar sua capacidade de detecção ao longo das épocas de treinamento.



Ao longo do treinamento, tanto o gerador quanto o discriminador tentam melhorar suas capacidades e competem entre si, como adversários. O gerador tenta criar imagens cada vez melhores para enganar o discriminador, e o discriminador tenta avaliar com mais exatidão a cada época. O treinamento acaba quando o discriminador não é mais capaz de distinguir entre as imagens falsas e reais.

Teste

Teste realizado com uma rede generativa treinada no dataset MNIST:



Construindo o Mini-GPT

Transformers

O vídeo apresentado por Andrej Karpathy -co-fundador da Open-AI- ensina a construir o modelo mais básico do ChatGPT a partir do artigo "Attention is all you need", publicado por pesquisadores da Google. O artigo introduz o *transformer*, uma arquitetura proposta para processamento de linguagem natural(NLP) que se baseia em um mecanismo chamado *self-attention*.

Self-Attention

O Self-Attention ou "Auto-atenção", relaciona diferentes posições de uma única sequência para computar uma representação da sequência, permitindo encontrar correlações entre as palavras de entrada e sua estrutura sintática, em modelos *Sequence to Sequence*.

- **Sequence-to-Sequence:** Um modelo que recebe uma sequencia de itens(palavras, letras, tokens etc) e devolve outra sequência de itens.

A autoatenção permite que cada token de entrada troque informações com todos os outros tokens da mesma sequência, independentemente de sua posição. Isso permite que o modelo capture relações semânticas complexas entre os tokens, superando as limitações que as *RNN's* tem devido ao tamanho da janela de contexto. Além disso, a autoatenção permite o processamento de todos os tokens em paralelo, o que acelera o treinamento.

Implementando o mini-gpt

O modelo desenvolvido no tutorial foi implementado com a biblioteca **PyTorch**, e treinado com o texto Tiny-Shakespeare. Ele foi definido como uma sequência de blocos de atenção seguidos por camadas lineares, fazendo as etapas:

1. Codifica os tokens em vetores de embedding.
2. Adiciona embeddings posicionais, que informam a posição dos tokens na sequência.
3. Passa pelos blocos Transformer, aplicando autoatenção e camadas FeedForward.
4. Retorna as previsões para o próximo caractere (usadas para calcular a perda com *cross_entropy*).

O modelo é treinado por meio de iterações, onde para cada iteração:

1. Um lote de dados é obtido.
2. O modelo faz previsões e a perda é calculada.
3. O otimizador ajusta os pesos do modelo com base no gradiente da perda.

Teste 1:

step 4999: train loss 1.6645, val loss 1.8286

KING RICHARD II:
Shal lifest made to bub, to take Our my dagatants:
Whith foul his vetward that a endrer, my fears' to zorm heavens,
Oof it heart my would but
With ensengmin latest in ov the doest not.

WARWICK:
Welll now, and thus quechiry: there's speak you love.
In Bodiet, and whom the sclittle
Enout-now what evily well most rive with is compon to the me
Town danter, If so;
Ange to shall do aleous, for dear?

KING HENRY VI:
A life, but a
ards bring Edward?

GROKE:
As is no Ruints I am you! who neet.
Pom mary thou contrantym so a theense.

QUEEN VINCENTIO:
O, sir, may in God't well ow, whom confessy.
Which migh.

ARCHILINIUS:
Dithul seaze Peed me: very it passce of's cruport;
How what make you fear tals: there loves
Tunkistren in deed, is xment.

Teste 2:

Hiperparâmetros alterados:

- **max_iters = 5000 -> 15000** : O aumento no número de épocas pode fazer com que o modelo convirja melhor.
step 14999: train loss 1.5217, val loss 1.7069

And they brinch'd, he son, but
bucket buble and Sthrives me?
I am such him foul hath be?
Fietle water with my feansure men our
Yount, to time come milind;
Who eseithers, I in latest in overse,
Be the news is wantoneldly die; let him, courity:
Auffinish he, yet let me nor
To tell deserveny: the cause
To Willow her evicy, when so rive with in his son
A life but to hingst firs son; if he must not male of my sad
Reed my offering, be!

GILLOUCESTER:
They louteman be to marchey ire to change! I had news noke married for this eymes had;
And to powerne the men.
Was stands bline one liegived, with
confisty to the mighty most wirthly in the seazent.
But invenes mount commons in meeting.

DUCHESS OF YORK:
He as neither looks, she
surpinicion one. Moly ashest! bear
that heaving with so! who buck now,
Whose lack, 'tis years' and subject.
My hands vaning now thief
Thou hat is is deterve no e subjectip'ding,
Puwitience to cousin England, madon sold;
For for me watch the
only Tigne; a;
you conto enemity, and we hasg'd I huse,
A whence outchan his constails; fall not wen or love
As be resenting in flears his breath officien of word,
Dosce to the
near wish I have a canith a polining one begoth.

O modelo teve uma melhora no resultado das perdas de treino e validação, mas o resultado textual ainda não faz qualquer sentido, com palavras inventadas.

Teste 3:

Hiperparâmetros alterados:

- **n_embd = 64 -> 128**: aumentar o embedding ajuda o modelo a aprender representações mais detalhadas dos tokens.
- **n_layer = 4 -> 6**: aumentar o número de camadas transformer faz com que o modelo capture

- relações mais complexas no texto.
- **block_size = 32 -> 64:** o aumento da janela de contexto faz com que o modelo considere mais tokens nas previsões, e isso pode ajudar a criar textos com algum sentido.
 - **batch_size = 16 -> 32:** aumento na quantidade de sequências processadas em paralelo, pode acelerar o treinamento.
 - **learning_rate = 1e-3 -> 2e-3:** seu aumento pode acelerar a convergência, mas há o risco de perder estabilidade.
 - **dropout = 0.0 -> 0.1:** necessário para reduzir o overfitting.
 - **max_iters = 15000 -> 5000:** redução nas épocas para compensar o custo computacional do aumento dos outros parâmetros.

```
step 4999: train loss 1.3543, val loss 1.5645
```

The nameal and whose compostion.

Lords:
To commandly, now. Heavent questing whereby I hope
For he side your servius cole my heart. He'nry have not
Have thousands to act as and earmiss.

First Nepare, more liege?

Clown:
Poor marry remembrard, what's I'fanch.

AUXETES:
Shall you would appear fronce! fare you to
less your leave ascure: gently thinst sensue,
Though On
Curse to pagues. Or cansels is he end,
Out, Lovesbot, and lies thee to pan the by cool.
Gallance then, the time of the thronv plafe,
And that I slain the strive, sort,
Should being my news incelity.

RICHARD OF AUM:
Exnaster more than born expect intent ,
Bear the heart of Rome.
No!

Os valores de perda de treino e validação reduziram ainda mais, e o texto parece mais estruturado e com menos palavras inventadas, mas ainda sem ter muito sentido.

Teste 4:

Daqui em diante apenas aumentei os parâmetros pelos mesmos motivos citados no último teste, para ver se há alguma melhora.

- **n_embd = 128 -> 128:**
- **n_layer = 8 -> 8:**
- **n_head = 4 -> 8:**
- **block_size = 64 -> 256:**
- **batch_size = 16 -> 32:**
- **learning_rate = 2e-3 -> 3e-3:**
- **dropout = 0.1 -> 0.2:**
- **max_iters = 5000 -> 5000:**

```
step 4999: train loss 1.2252, val loss 1.4798
```

Then in our healm that's many in lecestors?
O his vault to Lord Volscians
That he would, run usoa much and oractience,
Sevirgn at staater, must be night,
A manate our imprior born than abouther adve.

BUTON:
Why whenl's grumer in me belly long me,
Could smile the ounty with a churish,
Still her gentlemen?
Go, Warwick him ta'en her revenges to hear him.

CAMILLO:
Hither her hears; he mine, and coundert
He mercy uncles you, Force and for when yours;
O God seep hence! hear Euntua here in hell joyfull.

CLEOS:
Fait Lord, sir, you have heair all about me I unfly hund
And from you their ids my nderprise,
And from us I can past. Why be pasties,
Would have this coden, who comman'st cousin?

O texto melhorou, tendo mais sentido em algumas frases, mas ainda sem criar nenhum contexto. Os valores de perda de treino e validação também melhoraram, mas o modelo está em overfitting:

- val loss - train loss =
1. 1,8286-1,6645=0,1641
 2. 1,708-1,5217=0,1863
 3. 1,5645-1,3543=0,2102
 4. 1,4798-1,2252=0,2546

Teste 5:

- **n_embd = 128 -> 128:**
- **n_layer = 8 -> 8:**
- **n_head = 8 -> 8:**
- **block_size = 64 -> 256:**
- **batch_size = 64 -> 64:**
- **learning_rate = 3e-3 -> 4e-3:**
- **dropout = 0.2 -> 0.2:**
- **max_iters = 5000 -> 7000:**

```
step 6999: train loss 1.3796, val loss 1.6016
```

My present myself; be so Lovey'd knie
With lustand akish justice. How, neter diesh, as ado,
As I faill! cert's our fear which hate,
And neight heaver justs to bed himsouthe,
Are pair seem furse;
And qater to the many service;
But why you may that my see, is the newsy
Against up to hang of thy some goders: away!

FLARDY SURY:
Our thelest go their coporter often me.
But let us thy ad noless keellow the falser,
Even is my face, lame harded in bestrued, vantage
He greening credempt rehild prevounds to thee,
But but weep time, cause so lot uchiden;
As wee it an I chased thee be speaks -
That mething fortune, noisure, and my your here
Her all then he daybjoy. Dears I sir.

WARWICK:
Saint this hease you
The shame with twenth be so interchily enempts,
When thou by your hate hadst the hearting willy chese fried?

KING RICHARD II:
Say, sir. Hy, then is't state a doth,
The ceal embealt you which your mirning were one.

Com o aumento na taxa de aprendizado e no número de épocas, os valores de perda no treino e teste aumentaram, mesmo reduzindo o overfitting. não houveram melhoras significativas em comparação com o teste anterior.

- val loss - train loss =
1. 1,8286-1,6645=0,1641
 2. 1,708-1,5217=0,1863
 3. 1,5645-1,3543=0,2102
 4. 1,4798-1,2252=0,2546
 5. 1,6016-1,3796=0,222

Conclusão

As atividades desta etapa abordaram conceitos importantes do estado da arte das redes neurais generativas, como os **Variational Auto-Encoders** (VAEs) e as **Generative Adversarial Networks** (GANs), explorando como esses modelos geram dados realistas.

A implementação do **mini-GPT** utilizando **Transformers** introduziu o conteito de **self-attention** e sua capacidade em capturar relações complexas em sequências de texto, superando limitações de modelos anteriores. Os conhecimentos adquiridos são essenciais para o desenvolvimento de redes neurais mais avançadas e eficientes.