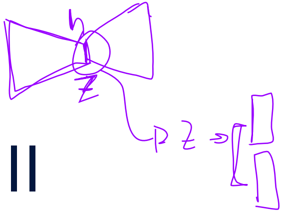
- Under/Over complete → # latents
  - ↳ $h < v$ ↳ $h > v$

- → MSE $\hat{x} \leftarrow x$
- → cross entr → bin



Autoencoders II

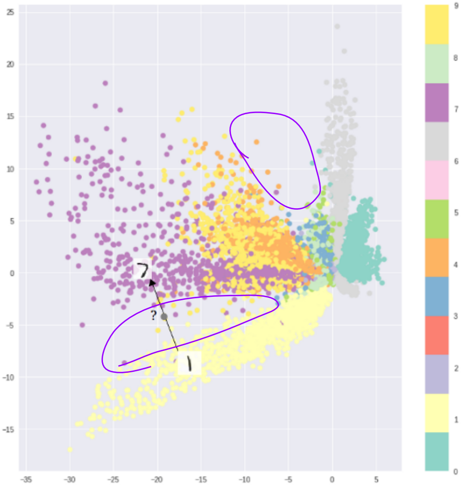
Advanced Institute for Artificial Intelligence – AI2

https://advancedinstitute.ai

Variational Autoencoders

Recapitulando: O problema com o autoencoder padrão

- Além de algumas aplicações eficientes como denoising autoencoders, eles são bastante limitados;
- O espaço latente para o qual convertem suas entradas e onde estão seus vetores de enconding pode não ser contínuo ou permitir fácil interpolação.
- Por exemplo, treinar um autoencoder no conjunto de dados MNIST e visualizar as codificações de um espaço latente 2D revela a formação de clusters distintos:



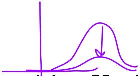
Recapitulando: O problema com o autoencoder padrão

- Ao construir um modelo generativo, não queremos replicar os dados de entrada:
  - Amostrar aleatoriamente do espaço latente, ou
  - Gerar variações em uma imagem de entrada, a partir de um espaço latente contínuo;
- Se o espaço tiver descontinuidades e você amostrar/gerar uma variação a partir daí, o decoder simplesmente gerará uma saída irreal;
  - O decoder não tem ideia de como lidar com aquela região do espaço latente;
  - Durante o treinamento, nunca viu vetores codificados vindos daquela região do espaço latente;

Definições

$Z \sim N$; B

- Os Variational Autoencoders (VAEs) têm uma propriedade fundamentalmente única que os separa dos autoencoders comuns:
  - Seus espaços latentes são, por desenvolvimento, contínuos;
  - A continuidade do espaço latente permite fácil amostragem aleatória e interpolação.
- Seu encoder não produz um vetor de codificação de tamanho n; $\rightarrow n/2$
- Em vez disso, ele gera dois vetores de tamanho n:
  - um vetor de médias, μ, e $\rightarrow N(\mu, \sigma)$
  - outro vetor de desvios padrão, σ.
  - A média e o desvio padrão da i-ésima variável aleatória, X_i da qual amostramos, para obter a codificação amostrada que passamos para o decodificador;



Definições

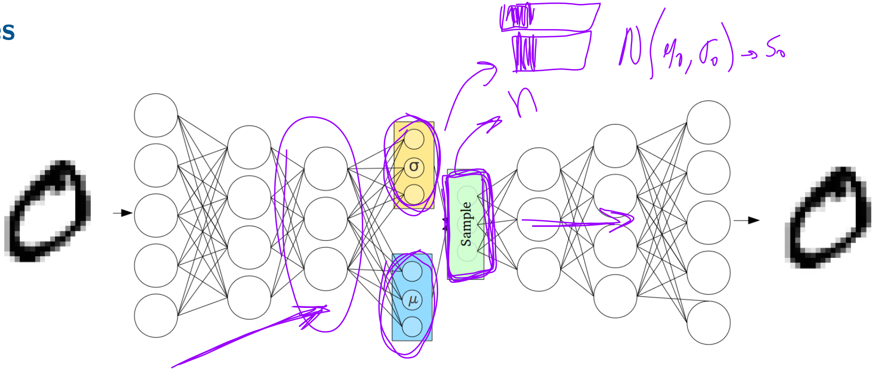


Figure: Variational Autoencoder com os vetores μ e σ.

Fonte: Variational Autoencoder architecture by Irhum

Example

No cenário onde temos um sinal de entrada com 500 características e pretendemos reduzir esse sinal para apenas 30, poderíamos pensar em construir um VAE da seguinte forma:

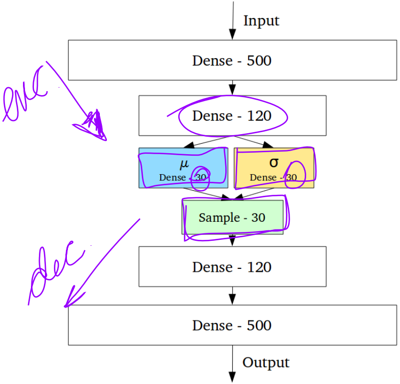


Figure: VAE que reduz as 500 dimensões de entrada para 30 dimensões no espaço latente.

Definições

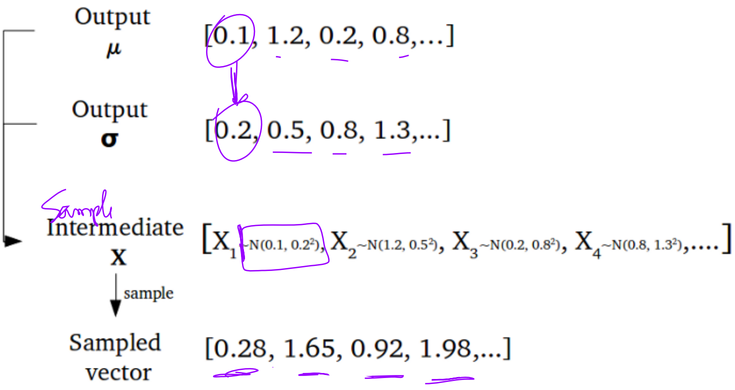


Figure: Como o passo forward funciona.

Variational Autoencoders

Definições

- Geração estocástica de vetores de codificação.
  - Para a mesma entrada, mantendo a média e o desvio padrão iguais, a codificação real irá variar em cada passagem devido à amostragem.
- O vetor da média controla onde a codificação de uma entrada deve ser centralizada;
- O desvio padrão controla quanto da média a codificação pode variar (a área)

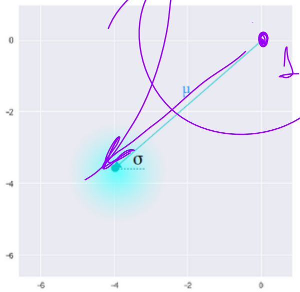


Figure: μ e σ para controlar a amostragem.

Definições

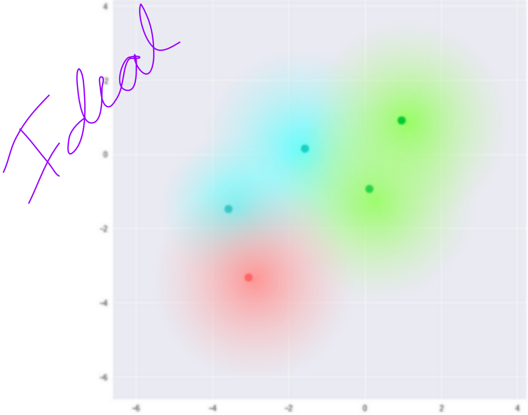
- Não apenas um único ponto no espaço latente se refere a uma amostra dessa classe.
- Todos os pontos próximos se referem ao mesmo em um raio-$sigma$;
- O objetivo aqui é tentar criar um espaço latente mais homogêneo, eliminando a descontinuidade;
  - O modelo está agora exposto a um certo grau de variação local variando o enconding de uma amostra;
  - Queremos sobreposição entre amostras que também não sejam muito semelhantes;
    - Interpolação entre classes;

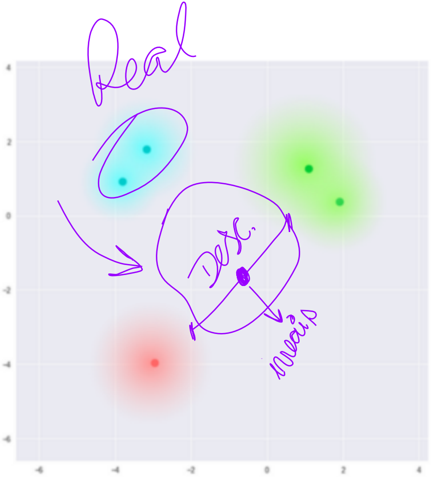
Definições

- Não há limites para os valores que os vetores μ e σ podem assumir:
  - Encoder pode aprender a gerar μ muito diferentes para diferentes classes, agrupando-as e minimizando σ
  - Pode chegar a um ponto que parece um único ponto.
- Desejável: Codificações que sejam o mais próximas possível e ainda assim distintas, permitindo uma interpolação suave e possibilitando a construção de novas amostras.

Definições

O que queremos e o que podemos obter:





Definições - A divergência KL

$KL(q||p) = -\sum q \log \frac{p}{q}$ 0

- Kullback-Leible divergence $KL \geq 0$
- Mede o quanto eles divergem entre si;
- Para VAEs, o custo pela KL é equivalente à soma de todas as divergências KL entre a componente $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ e a normal padrão.
  - Essa medida é minimizada quando $\mu_i = 0$ e $\sigma_i = 0$
- Quando a divergência é calculada entre distribuições univariadas, ela pode ser simplificada para [1]:

$$\sum_{i=1}^{n} \sigma_i^2 + \mu_i^2 - log(\sigma_i^2) - 1$$

[1] Deriving the KL divergence loss for VAEs

Definições - A divergência KL

- Essa perda força o encoder a distribuir todas as codificações uniformemente ao redor do centro do espaço latente;
- Usar puramente o resultado da loss KL resulta em um espaço latente com codificações densamente colocadas aleatoriamente, perto do centro do espaço latente;
- O decoder acha impossível decodificar qualquer coisa significativa deste espaço;

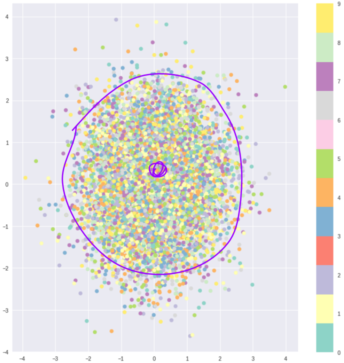


Figure: Espaço latente produzido por um VAE treinado apenas com a loss KL.

Variational Autoencoders

Agrupando as informações...

- Utilizar a divergência KL como mecanismo de penalização;
- Otimizar a loss composta (por exemplo reconstrução, ou entropia-cruzada) e a divergência KL;
  - Geração de um espaço latente que mantém a semelhança das codificações próximas;
  - Globalmente, é densamente compactado perto da origem do espaço latente;
  - Equilíbrio alcançado pela natureza formadora de agrupamentos da loss de reconstrução e pela natureza de empacotamento denso da loss KL;

mSE

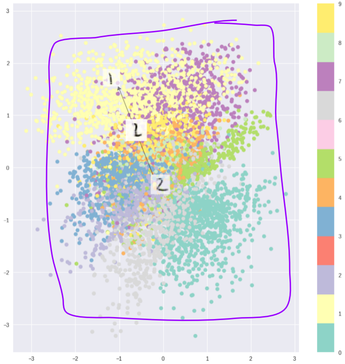
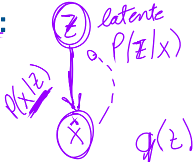


Figure: Empregando a loss combinada.

Um pouco da matemática dos VAE:

- → $H = -\sum p(x) \log P(x)$
- → $KL(q\|p) = -\sum q \log \frac{p}{q}$



$$\boxed{P(z|x) = \frac{P(x|z)\,P(z)}{P(x)} = \frac{P(x,z)}{P(x)}} \quad (1)$$

→ Markov . MCMC, Gibbs

→ VI → KL

$q(z) \approx P(z|x)$

(2) $KL(q(z)\|p(z|x)) = -\sum q(z) \log \frac{P(z|x)}{q(z)}$; Subst 1

$$= -\sum_z q(z) \log \frac{P(x,z)}{P(x)\,q(z)} = -\sum q(z)\Big[\underbrace{\log \frac{P(x,z)}{q(z)} - \log P(x)}_{+\log \frac{1}{P(x)}}\Big]$$

$$\Rightarrow KL = -\sum q(z) \log \frac{P(x,z)}{q(z)} + \sum q(z) \log P(x) \longrightarrow$$

Um pouco da matemática dos VAE:

$$KL = -\sum q(z)\log\frac{P(x,z)}{q(z)} + \sum q(z)\log P(x) = -\sum q(z)\log\frac{P(x,z)}{q(z)} + \log P(x)$$

$\sum q(z) = 1$ dist Prob

$$\frac{P(x|z)P(z)}{P(x)} = \frac{P(x,z)}{P(x)}$$

$$\to P(x,z) = P(x|z)P(z) \quad (4)$$

$$\log P(x) = -\sum q(z)\log\frac{P(x,z)}{q(z)} + KL(q(z)\|p(z|x)) \geq 0 \quad (3)$$

max

$\mathcal{L}$

min

$-KL(q(z)\|p(z))$

KL

$E_q \log P(x|z)$

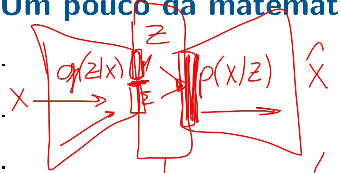
MSE

$$4 \to \mathcal{L} = +\sum q(z)\log\frac{P(x|z)P(z)}{q(z)} \Rightarrow \mathcal{L} = \sum q(z)\left[\log\frac{P(x|z)}{q(z)} + \log P(z)\right] \Rightarrow \mathcal{L} = \sum q(z)\log\frac{P(z)}{q(z)} + \sum q(z)\log P(x|z)$$

Variational Autoencoder

Um pouco da matemática dos VAE:

$$q(z|x) = \frac{q(x|z)\,q(z)}{q(x)} \rightarrow \boxed{q(z) = \frac{q(z|x)\,q(x)}{q(x|z)}}$$



$$E_{q\sim z} \log p(\hat{x}|z) = E_q \log\left(e^{|x-\hat{x}|^2}\right)$$

$$E_q |x-\hat{x}|^2$$

$$z \sim N(\mu, \sigma) \rightarrow N(\mu, \Sigma)$$

$$z \sim N \rightarrow e^{|x-\hat{x}|^2} \rightarrow \boxed{|x-\hat{x}|^2}$$

Reparam. trick

$$z = \mu + \sigma \cdot N(0,1)$$

Variational Autoencoder

Um pouco da matemática dos VAE:

-
-
-
-
-
-
-
-

Um pouco da matemática dos VAE:

Um pouco da matemática dos VAE:

-
-
-
-
-
-
-
-

Um pouco da matemática dos VAE:

Um pouco da matemática dos VAE:

-
-
-
-
-
-
-
-