

Auto-supervisão e Aprendizado Contrastivo

Redes Neurais e Aprendizado Profundo

Moacir A. Ponti

`www.icmc.usp.br/~moacir` — `moacir@icmc.usp.br`

São Carlos-SP/Brasil

Agenda

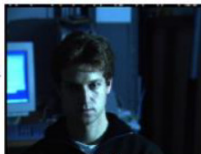
Aprendizado Contrastivo

Aprendizado auto-supervisionado

Lidando com variações intra-classe



1.50



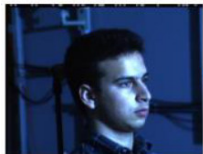
2.90



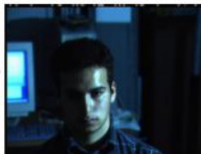
2.13



2.26



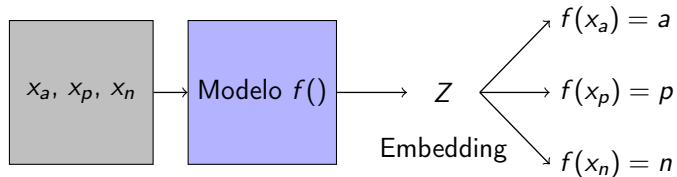
4.10



Lidando com variações intra-classe

- ▶ Aprender a partir de instâncias diretamente para a saída pode tornar o modelo dependente de características que não representam o que gostaríamos
- ▶ A saída: aprender a partir de grupos de exemplos, em particular pares ou triplas

FaceNet / Triplet loss



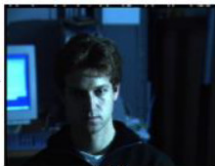
$$\|a - p\|^2 - \|a - n\|^2$$

- O objetivo é aprender representação que obedeça distâncias

Lidando com variações intra-classe



1.22



1.33

1.04

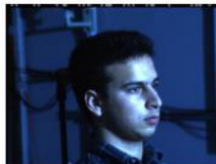


1.33

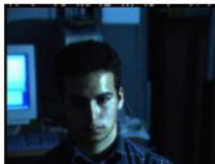


1.26

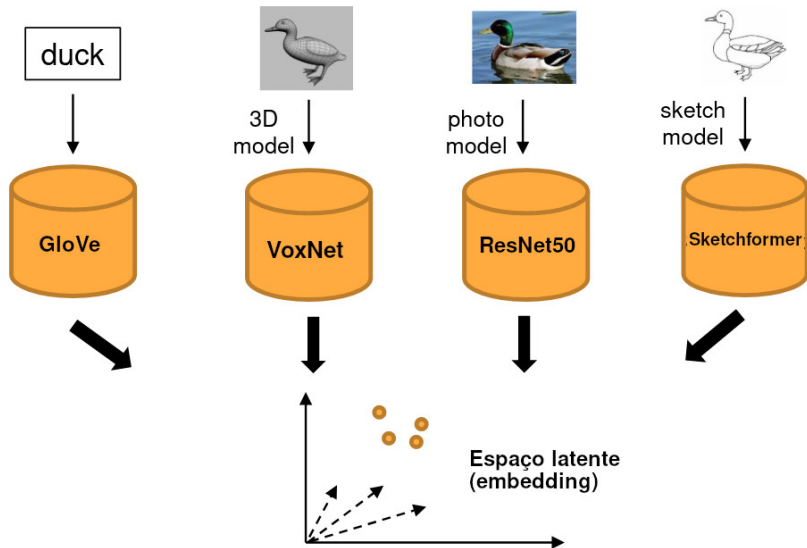
0.78



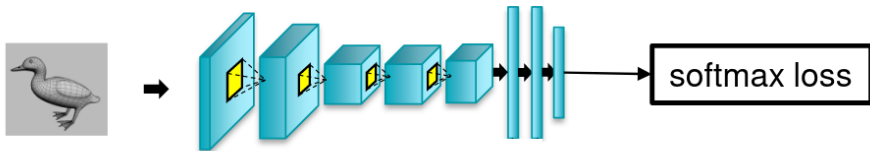
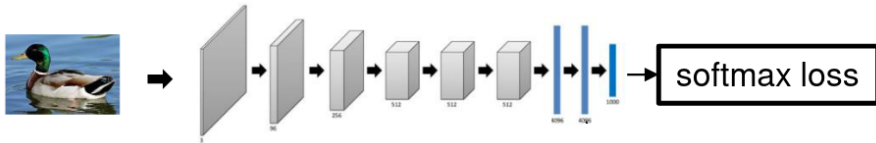
0.99



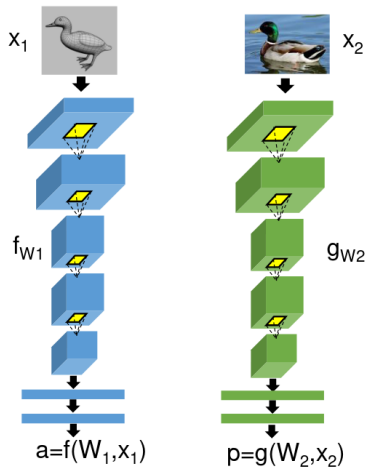
Redes multi-fluxo e aprendizado multimodal



Redes multi-fluxo e aprendizado multimodal



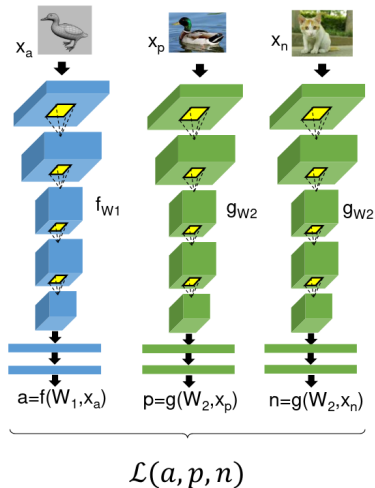
Redes com função contrastiva



- ▶ *Entrada:* par de exemplos x_1, x_2
- ▶ Modelos podem ser os mesmos ou diferentes (depende dos domínios)
- ▶ Função de custo considera as representações a, p obtidas da saída de uma das camadas
- ▶ Se p é positivo, então $y = 0$, senão $y = 1$, cancelando sempre um dos termos

$$L(a, p) = \frac{1}{2}(1 - y)|a - p|^2 + \frac{1}{2}y[\max(0, m - |a - p|)^2]$$

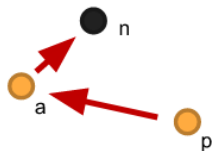
Redes triplet



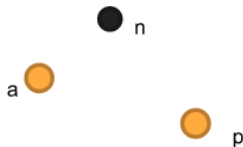
- ▶ *Entrada:* tripla x_a, x_p, x_n
- ▶ Modelos podem ser os mesmos ou diferentes (depende dos domínios)
- ▶ Função de custo considera as representações obtidas da saída de uma das camadas: a, p, n

$$L(a, p, n) = \frac{1}{2} [\max(0, m + |a - p|^2 - |a - n|^2)]$$

Intuição das funções de custo

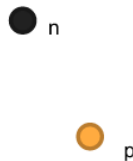


Before training



Contrastive loss

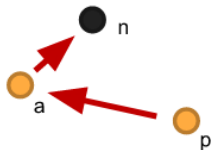
$$L(a, p) = \frac{1}{2} (1 - y) |a - p|_2^2 + \frac{1}{2} y \{ \max(0, m - |a - p|_2^2) \}$$



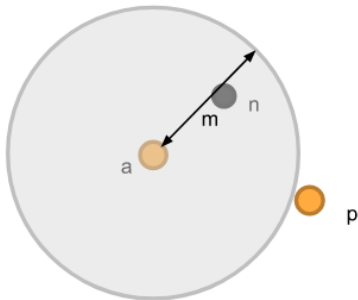
Triplet loss

$$L(a, p, n) = \frac{1}{2} \{ \max(0, m + |a - p|_2^2 - |a - n|_2^2) \}$$

Intuição das funções de custo

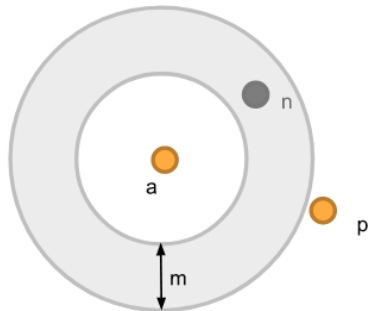


Before training



Contrastive loss

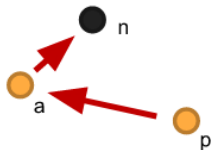
$$L(a, p) = \frac{1}{2} (1 - y) |a - p|_2^2 + \frac{1}{2} y \{ \max(0, m - |a - p|_2^2) \}$$



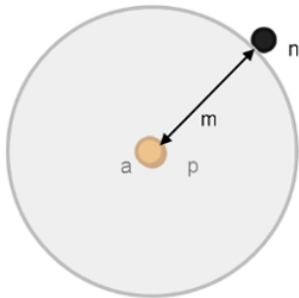
Triplet loss

$$L(a, p, n) = \frac{1}{2} \{ \max(0, m + |a - p|_2^2 - |a - n|_2^2) \}$$

Intuição das funções de custo

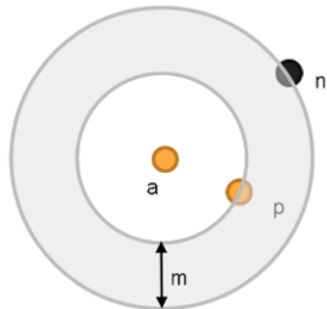


Before training



Contrastive loss

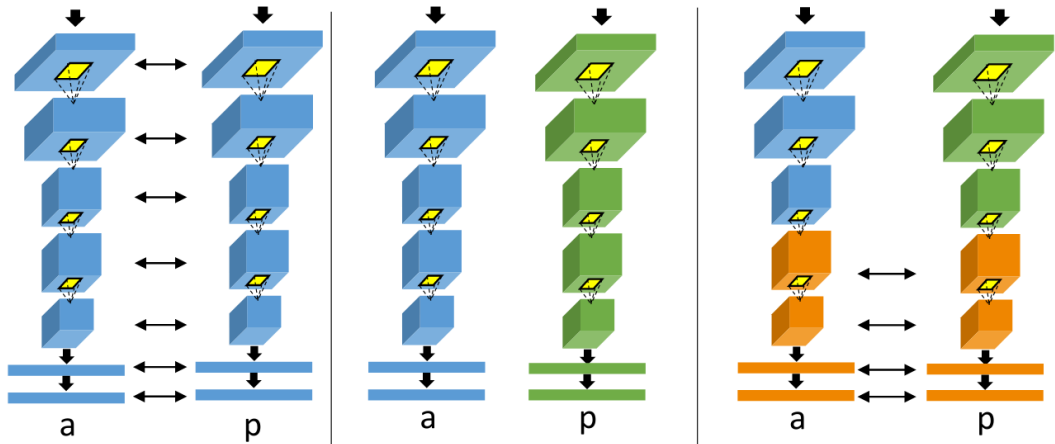
$$L(a, p) = \frac{1}{2} (1 - y) |a - p|_2^2 + \frac{1}{2} y \{ \max(0, m - |a - p|_2^2) \}$$



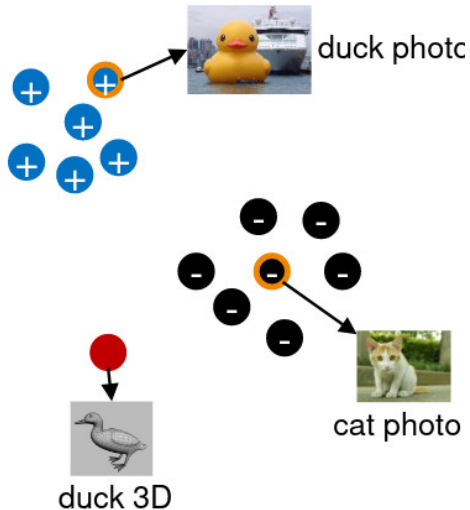
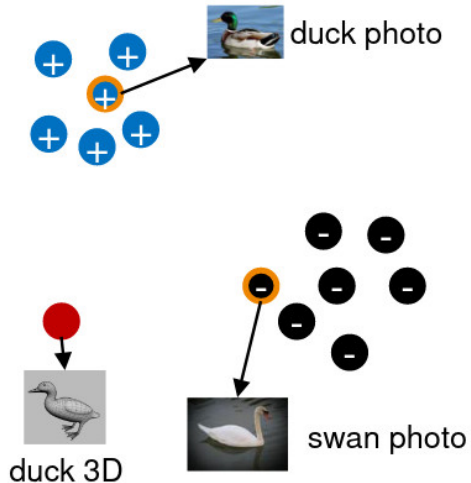
Triplet loss

$$L(a, p, n) = \frac{1}{2} \{ \max(0, m + |a - p|_2^2 - |a - n|_2^2) \}$$

Compartilhamento de pesos



Estratégia de treinamento: hard positive/negative



Agenda

Aprendizado Contrastivo

Aprendizado auto-supervisionado

Revisitando categorias de aprendizado

Aprendizado por reforço

- ▶ retorno fraco a cada etapa
- ▶ funciona bem quando episódios são fáceis de computar/simular

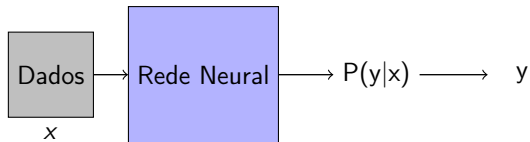
Aprendizado supervisionado

- ▶ retorno a cada etapa depende da variabilidade e quantidade de dados
- ▶ mas raramente há dados abundantes

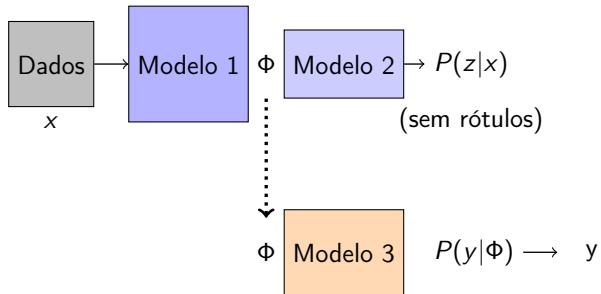
Aprendizado auto-supervisionado

- ▶ retorno a cada etapa é similar ao supervisionado, mas computado a partir dos dados de entrada
- ▶ podemos gerar número enorme de dados para treinamento

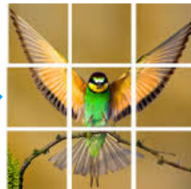
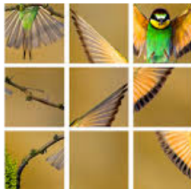
Aprendizado supervisionado para auto-supervisionado



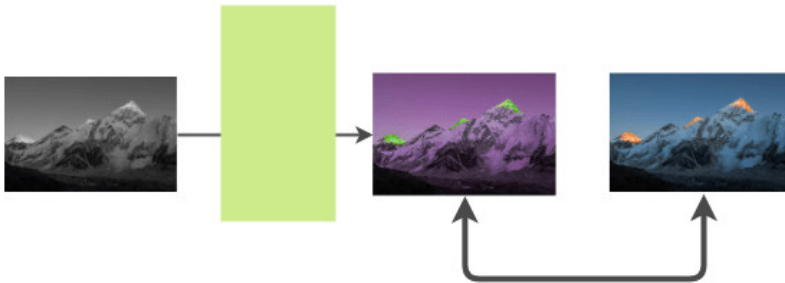
Aprendizado supervisionado para auto-supervisionado



Rótulos computáveis: rotação e quebra-cabeça



Tarefas computáveis: colorização



Tarefas auxiliares: preencher lacunas

Deixa o menino _____ bola e __pren_____

Tarefas auxiliares: preencher lacunas

Deixa o menino pegar bola e aprender

Outras tarefas possíveis

- ▶ Redes geradoras
- ▶ Denoising Autoencoders
- ▶ Pseudo-labels com agrupamento
- ▶ Aprendizado contrastivo multidomínio: áudio + imagem, áudio + texto

Referências: Redes multi-fluxo e aprendizado de métricas

- ▶ Blog com conteúdo sobre redes siamesas (multi-fluxo):
<https://towardsdatascience.com/illustrated-guide-to-siamese-network-3939da1b0c9d>
- ▶ Blog com conteúdo sobre aprendizado de métricas:
<https://towardsdatascience.com/the-why-and-the-how-of-deep-metric-learning-e70e16e199c0>
- ▶ Artigo survey na área:
https://www.researchgate.net/publication/335314481_Deep_Metric_Learning_A_Survey
- ▶ Notebook com exemplo de Contrastive Learning:
<https://colab.research.google.com/github/keras-team/keras-io/blob/master/examples/vision/ipynb/supervised-contrastive-learning.ipynb>

Referências: Aprendizado auto-supervisionado

- ▶ Lista de recursos: <https://github.com/jason718/awesome-self-supervised-learning>
- ▶ Artigo "self-taught learning" (2007): <https://ai.stanford.edu/~hllee/icml07-selftaughtlearning.pdf>
- ▶ Artigo "How Well Do Self-Supervised Models Transfer?" (2021): <https://arxiv.org/pdf/2011.13377>
- ▶ Artigo "Predicting What You Already Know Helps" (2021): <https://arxiv.org/pdf/2008.01064.pdf>
- ▶ Palestra do Yan LeCun <https://www.youtube.com/watch?v=SaJL4SLfrcY>
- ▶ Notebook com exemplo de Self-supervised contrastive learning: <https://colab.research.google.com/github/keras-team/keras-io/blob/master/examples/vision/ipynb/simsiam.ipynb>