



Uma Abordagem de Aprendizado Profundo e *Data Augmentation* para Classificação de Imagens Aéreas

Willian Dihanster Gomes de Oliveira, Lilian Berton
Aprendizado de Máquina e Reconhecimento de Padrões - Prof^a Dra. Ana Carolina Lorena
Instituto de Ciência e Tecnologia, Universidade Federal de São Paulo
29 de Junho de 2018
São José dos Campos - SP

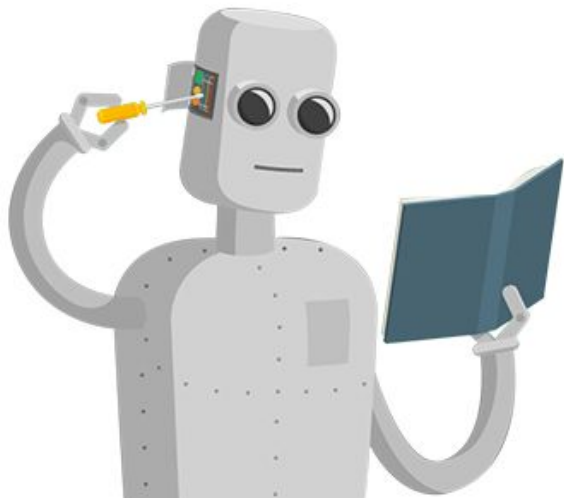


Uma Abordagem de Aprendizado Profundo e *Data Augmentation* para Classificação de Imagens Aéreas

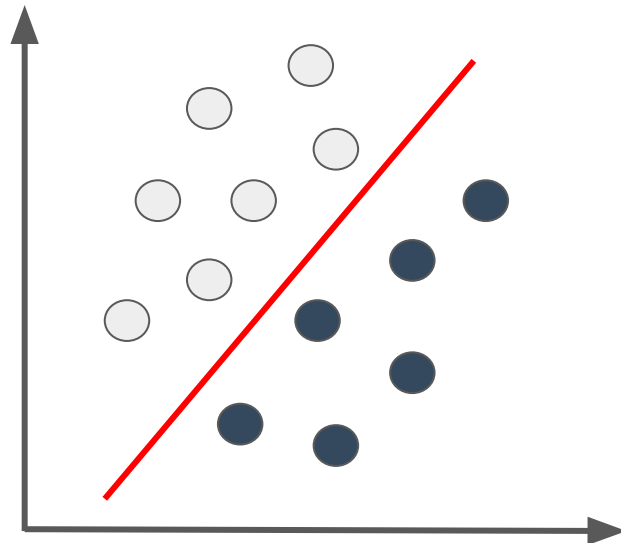
Willian Dihanster Gomes de Oliveira, Lilian Berton
Inteligência Artificial - Profº Dr. Vinicius Veloso Melo
Instituto de Ciência e Tecnologia, Universidade Federal de São Paulo
30 de Junho de 2018
São José dos Campos - SP

Aprendizado de Máquina e Classificação

- Aprendizado de Máquina (AM):
 - Aprender a partir de exemplos e melhorar automaticamente com a experiência. [1]
- Classificação: Uma das maiores aplicações do AM. Pode automatizar tarefas que necessitem reconhecimento automático.
- Ex: Classificação de Imagens aéreas de plantações de café, ajudar agricultores a identificar localizações de sua plantação que possuem um pé de café.

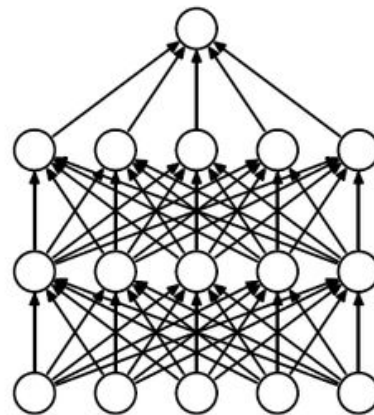
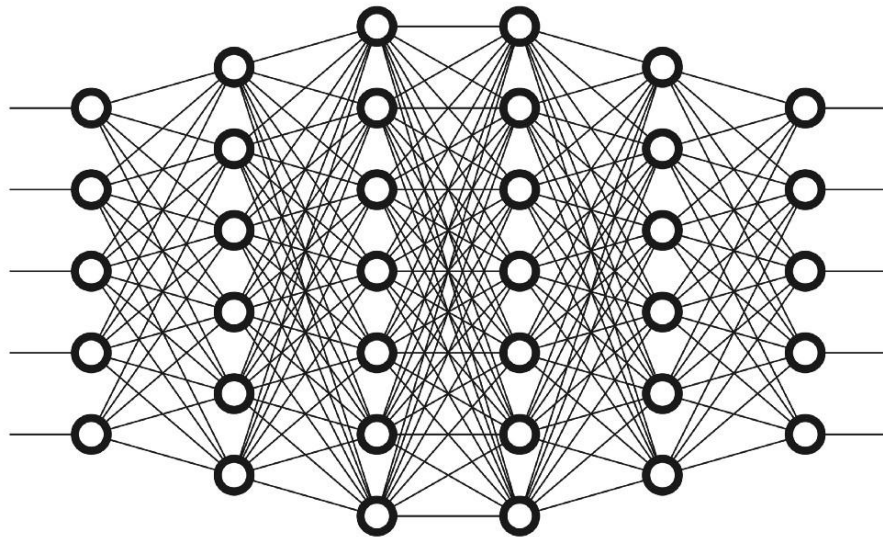


Robô ajustando seus parafusos para aprender. Disponível em:
<http://fiori-implementation.com/blog/make-clean-breast-machine-learning/>

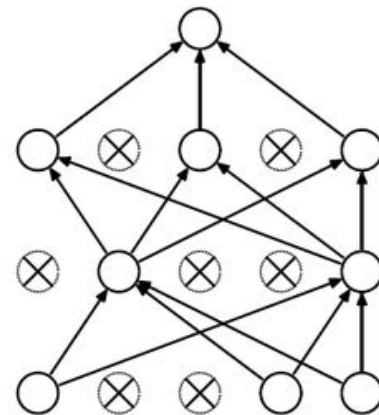


Deep Learning

- Um classificador em destaque são as redes neurais.
- *Deep Learning*: Redes Neurais Profundas -> Maior capacidade de abstração. [2]
- Alguns modelos:
 - *Multilayer Perceptron*.
 - *Multilayer Perceptron com Dropout*. [3]



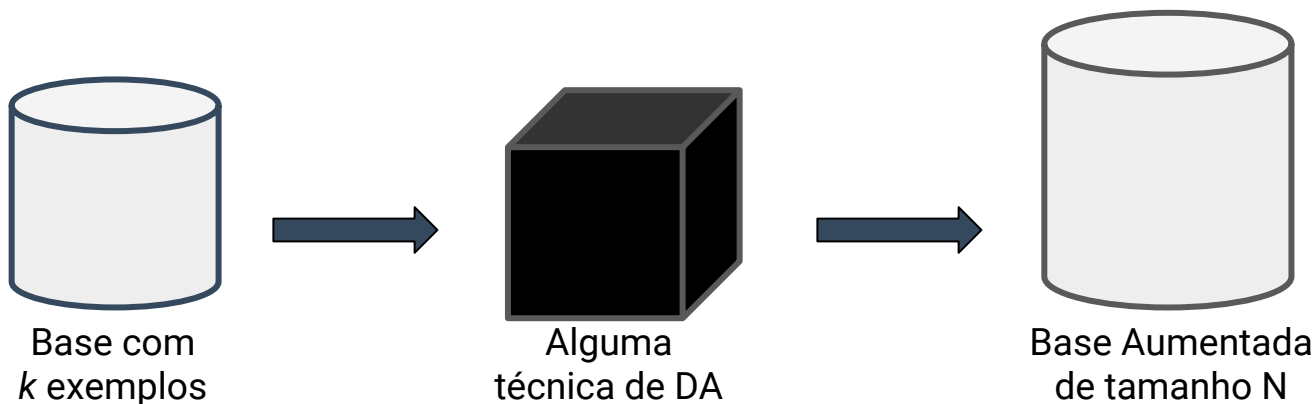
(a) Standard Neural Net



(b) After applying dropout.

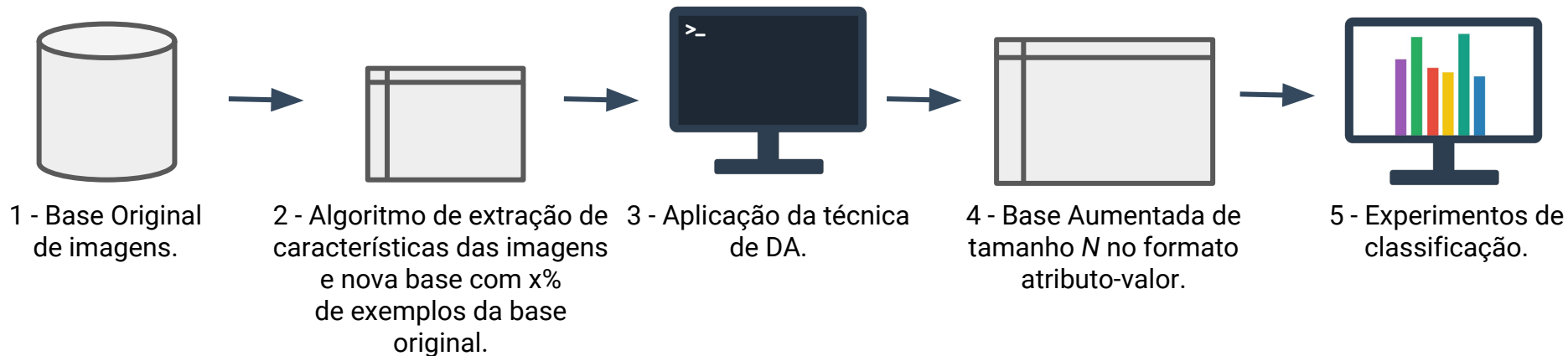
Data Augmentation (DA)

- Aprendizado Supervisionado: quanto mais dados (rotulados), maior a acurácia.
- Mas, dados rotulados são custosos e difíceis de se obter.
- *Data Augmentation*:
 - Técnicas computacionais para aumentar bases de dados.
- Objetivo: estudo de duas técnicas de DA para a classificação de imagens.



Metodologia

- A metodologia é descrita como a sequência de passos a seguir:

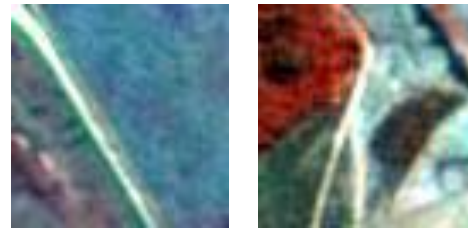


Base de Dados

- A base de dados utilizada (*Brazilian Coffee Scenes Dataset* [4]) é composta por imagens de satélite de plantações de café, em cidades de Minas Gerais pelo sensor SPOT.
- Cada imagem possui 64x64 pixels e foram classificadas manualmente por pesquisadores agrícolas como "coffee" e "non coffee".
- 2876 imagens, 50% de coffee e 50% non coffee.



Exemplos rotulados como coffee.



Exemplos rotulados como non coffee.

Técnica de DA 1: Transformação de Imagens (Imagens)

- Nessa abordagem, são escolhidas algumas imagens aleatoriamente do conjunto original e aplica-se alguns efeitos como corte, girar, *blur*, até expandir para um novo conjunto de tamanho N .



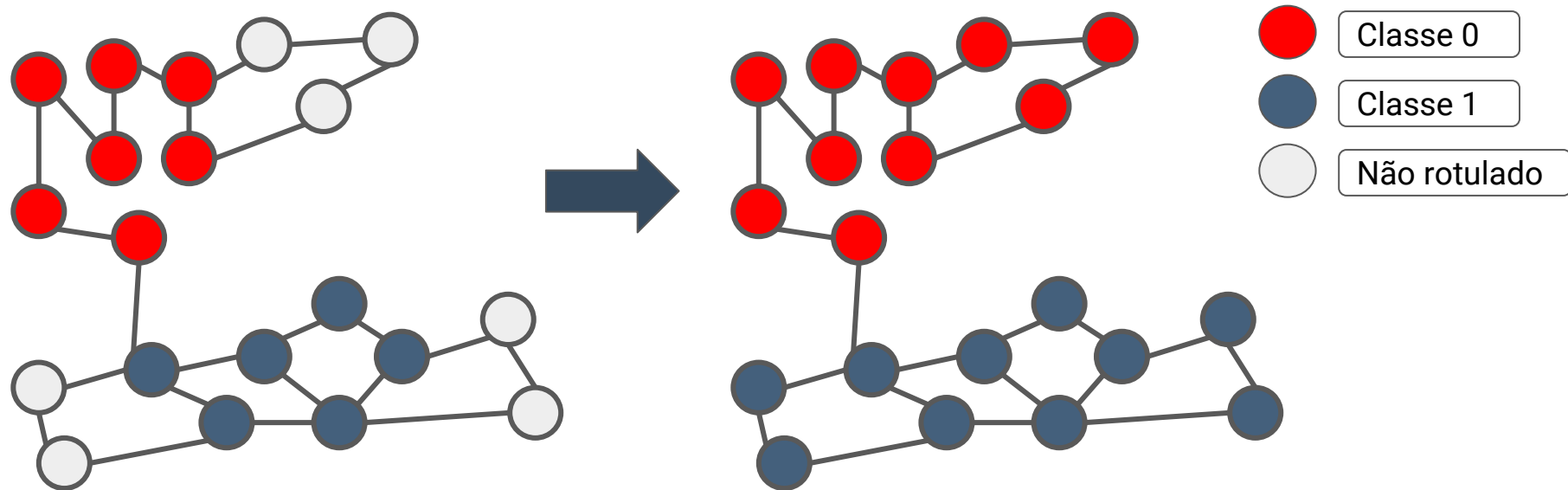
Exemplo aleatório
rotulado como coffee.



Exemplo de coffee depois de
aplicado efeito *blur* e girar.

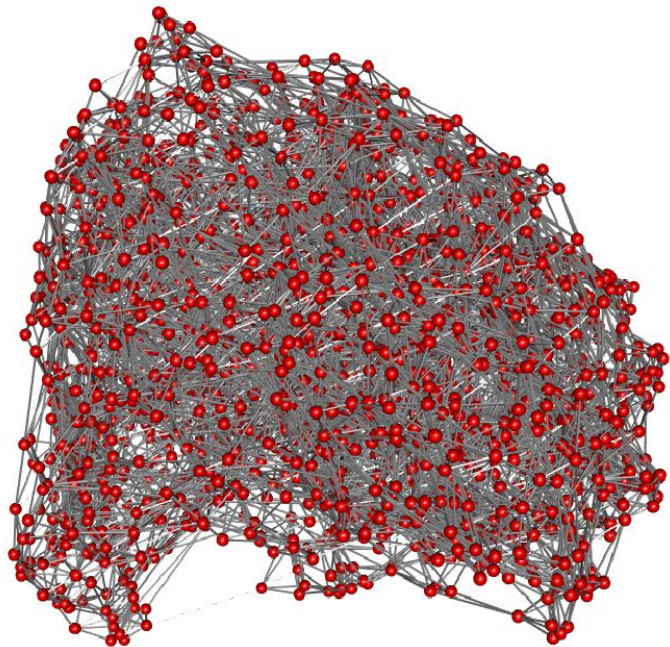
Técnica de DA 2: Aprendizado Semissupervisionado Baseado em Grafos

- Elementos rotulados e não rotulados. Pode-se tentar classificar esses dados não rotulados.
- Algoritmos de Propagação de Rótulos:
 - A partir de dados rotulados, propagar sua informação de classe para os dados ainda não rotulados.
 - Para isso pode-se usar um grafo.



Aprendizado Semissupervisionado Baseado em Grafos (Grafos-SSL)

- Constrói-se um grafo com a informação de semelhança dos exemplos e é aplicado o algoritmo de propagação de rótulos *Local and Global Consistency (LGC)* [5].
- Pontos próximos (local) e de mesma estrutura (global) são prováveis de terem mesmo rótulo.



Grafo construído para a propagação.

Algoritmo 1: Local and Global Consistency

Entrada: G um grafo ponderado $G(V, E, W)$.

Saída: Rótulos para o conjunto Y_u , $l + 1 \leq u \leq n$

início

Calcula-se a matriz de afinidade W , utilizando a seguinte fórmula:

$$W_{ij} = \begin{cases} \exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2}) & \text{se } i \neq j \\ 0 & \text{se } i = j \end{cases} \quad (19)$$

Calcula-se a matriz diagonal D , como: $D_{ii} \leftarrow \sum_j W_{ij}$.

Calcula-se a matriz laplaciana normalizada utilizando: $S = D^{1/2} W D^{1/2}$.

Escolhe-se um parâmetro α no intervalo $(0, 1)$.

repita

$F(t + 1) = \alpha S F(t) + (1 - \alpha) Y$.

até que F atinja a convergência;

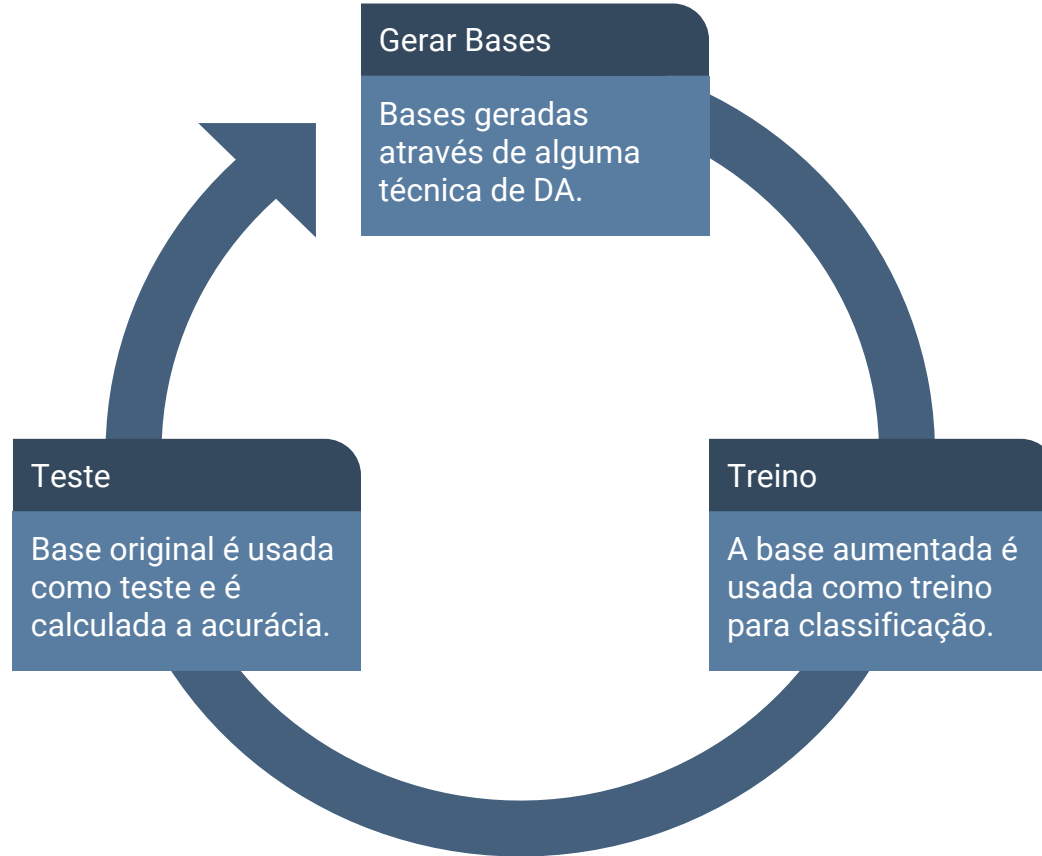
Ao final, obtém-se F^* .

Rotula-se as amostras x_i , com $y_i = \arg \max_{j \leq c} F_{ij}^*$.

fim

Algoritmo LGC.

Experimentos



Softwares e classificadores utilizados para os experimentos

- Softwares e bibliotecas utilizadas:
 - WEKA
 - TensorFlow para Python.
- Classificadores/Algoritmos utilizados (parâmetros padrão):
 - *Multilayer Perceptron*.
 - *Multilayer Perceptron* com *Dropout*.

Experimentos Iniciais

- Gerou-se bases de dados para cada técnica de DA considerada. Considerando 10% dos dados rotulados originais.
- Essas bases passaram por uma sequência de experimentos e escolheu-se a base (de cada técnica) com maior desempenho médio.

	Média	Desvio Padrão	Máximo	Mínimo
Imagens	78,66%	1,64	82,68%	70,97%
Grafos-SSL	80,04%	1,94	84,98%	73,61%
Original (10-fold)	84,37%	2,65	87,41%	79,73%

Variação de Parâmetros das RN

- Considerando 4 redes:
 - Rede 1: WEKA com 3 Camadas Ocultas.
 - Rede 2: TensorFlow com 3 Camadas Ocultas.
 - Rede 3: TensorFlow com 4 Camadas Ocultas (com *Dropout*).
 - Rede 4: TensorFlow com 5 Camadas Ocultas.
- Parâmetros Padrão (Principais) :
 - *Learning Rate* = 0.3.
 - *Training Time/EPOCHS* = 500.
 - 64 neurônios em cada camada oculta.
- Para os experimentos, varia-se um parâmetro, enquanto que os outros são mantidos fixo.

Learning Rate

- Valores considerados (0.03, 0.05, 0.1, 0.2, 0.3)
- Melhores resultados:
 - Imagens: 84.28% com a Rede 1 (WEKA 3 camadas)
 - Grafos-SSL: 83.83% com a Rede 3 (TensorFlow 4 camadas, com *Dropout*).

Valor	Rede 1	Rede 2	Rede 3	Rede 4
0.03	83.71	80.38	83.55	81.09
0.05	83.72	80.31	83.81	80.61
0.1	83.90	80.35	83.67	80.49
0.2	83.73	80.05	83.86	80.44
0.3	84.28	79.74	83.97	80.38

Variação do Learning Rate para Base DA Imagens.

Valor	Rede 1	Rede 2	Rede 3	Rede 4
0.03	81.47	82.89	83.83	82.98
0.05	83.55	82.88	83.77	82.86
0.1	83.73	82.72	83.80	82.99
0.2	83.38	82.85	83.77	82.68
0.3	82.61	82.68	83.70	82.80

Variação do Learning Rate para Base DA Grafos-SSL.

Training Time/Epochs

- Valores considerados (40, 100, 200, 400, 500, 600)
- Melhores resultados:
 - Imagens: 84.46% com a Rede 1 (WEKA 3 camadas)
 - Grafos-SSL: 84.11% com a Rede 1 (WEKA 3 camadas).

Valor	Rede 1	Rede 2	Rede 3	Rede 4
40	84.46	83.12	84.02	73.60
100	84.32	82.03	84.14	82.51
200	84.28	81.06	84.05	81.45
400	84.30	80.29	83.46	80.67
500	84.27	79.74	83.97	80.38
600	84.25	79.01	84.29	80.00

Variação do *Training Time* para Base DA Imagens.

Valor	Rede 1	Rede 2	Rede 3	Rede 4
40	82.78	82.30	83.33	82.25
100	82.79	82.62	83.92	82.78
200	82.37	82.94	83.74	82.94
400	83.38	83.07	83.75	82.98
500	82.64	83.09	83.81	83.14
600	84.11	82.97	83.78	83.18

Variação do *Training Time* para Base DA Grafos-SSL.

Número de Neurônios em cada camada oculta

- Valores considerados (32, 64, 128, 256 e 512)
- Melhores resultados:
 - Imagens: 84.28% com a Rede 1 (WEKA 3 camadas)
 - Grafos-SSL: 83.82% com a Rede 3 (TensorFlow 4 com *Dropout*).

Valor	Rede 1	Rede 2	Rede 3	Rede 4
32	84.27	80.16	84.18	80.49
64	84.28	79.74	83.80	80.27
128	83.94	80.06	83.80	80.27
256	84.11	80.24	83.98	80.16
512	84.01	80.15	83.80	80.04

Variação do Número de Neurônios para Base DA Imagens.

Valor	Rede 1	Rede 2	Rede 3	Rede 4
32	83.31	83.00	83.71	83.08
64	82.61	83.09	83.82	83.14
128	82.79	82.83	83.81	83.03
256	81.43	82.91	83.80	82.89
512	83.76	83.20	83.79	76.49

Variação do Número de Neurônios para Base DA Grafos-SSL.

Teste de Hipótese Paired t-test

- Será que alguma técnica de saiu melhor que a outra ?
 - Testes utilizando a Rede 4 (mais camadas ocultas) com *Learning Rate* = 0.3, *Training Time* = 500 e 64 neurônios em cada camada.
- Com esse teste, foi possível concluir que há uma diferença significativa das médias. Ou seja, a abordagem de Grafos se saiu melhor que a abordagem de Imagens na rede 4.
- Sendo assim, quando há a possibilidade de escolha entre usar uma das duas técnicas, pode haver uma preferência para a abordagem de grafos.

Medida	Base DA Imagens	Base DA Grafos
Média	80.30	82.91
Mediana	80.26	82.94
Desvio Padrão	0.2769	0.2758
Variância	0.0782	0.0741

Conclusões

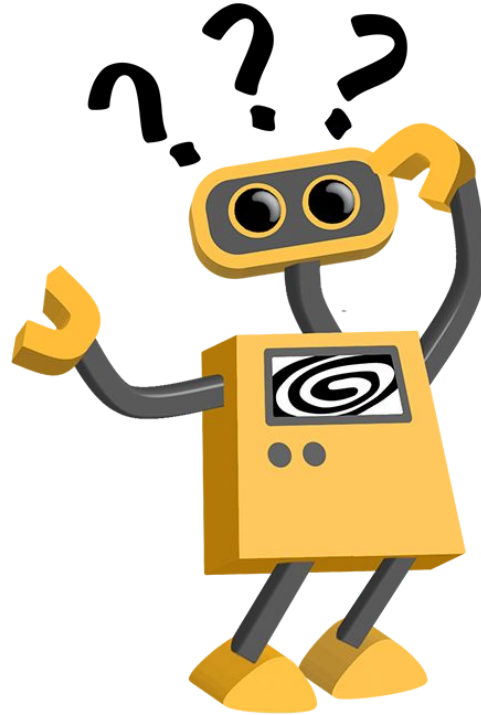
- Ambas técnicas podem ser consideradas eficientes para a aplicação escolhida. E que DA pode melhorar a acurácia no caso de poucos exemplos rotulados, além de ajudar na automatização de rótulos.
- *Deep Learning* pode melhorar acurácia em relação à técnicas comuns de classificação. (Imagens 82.58% para 84.46% e Grafos 83.17% para 84.11%)
- Técnica de *Dropout* pode ser utilizadas para melhorar a acurácia em *Deep Learning*.
- Grafos-SSL
 - Vantagens: Amplo, pode se aplicar diversos algoritmos de propagação, alterações na construção do grafo. Teste de Hipótese revelou melhor resultado para Rede de 5 camadas.
 - Desvantagens: Somente em abordagens em que se possuem elementos rotulados e não rotulados.
- Transformação de Imagens:
 - Vantagens: Pode-ser aplicado a qualquer conjunto de dados de imagens.
 - Desvantagens: Cuidado com a redundância ou perda de sentido nos dados.

Referências

- [1] Michalski, Ryszard S., et al., eds. *Machine learning: A multistrategy approach*. Vol. 4. Morgan Kaufmann, 1994.
- [2] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
- [3] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.
- [4] Penatti, Otávio AB, Keiller Nogueira, and Jefersson A. dos Santos. "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?." *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2015 IEEE Conference on. IEEE, 2015.
- [5] Zhou, Denny, et al. "Learning with local and global consistency." *Advances in neural information processing systems*. 2004.

Alguma Dúvida ?

- Obrigado pela atenção!



Disponível em: <https://fcit.usf.edu/matrix/project/robot-10-robot-with-questions/>