



Classificação de Imagens

Análise de Técnicas Recentes

João Brito, m9984
Ricardo Domingos, m10259

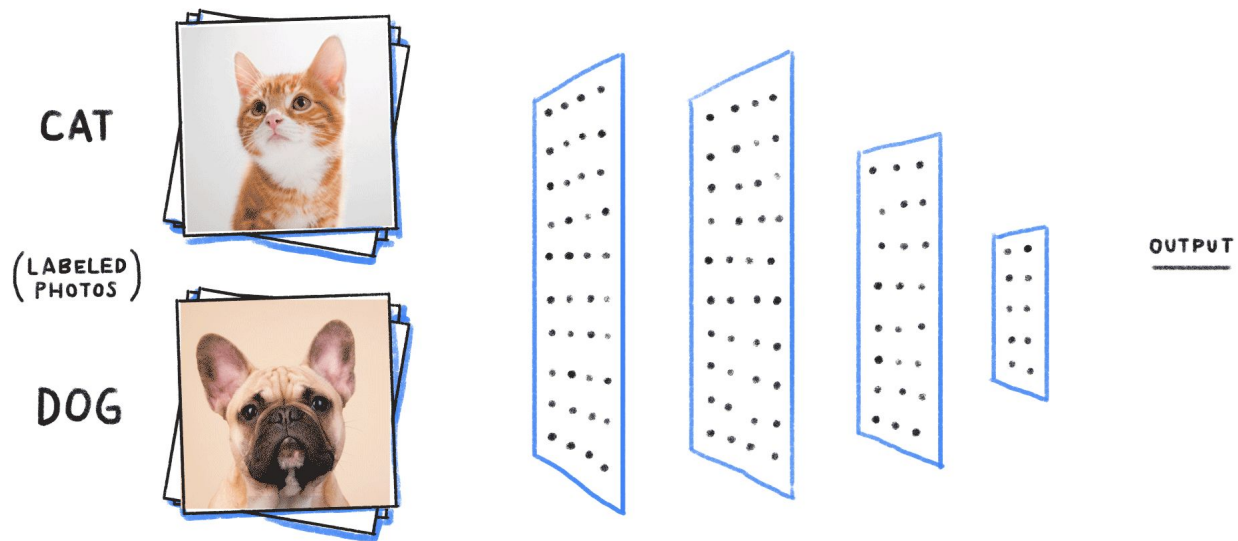
Inteligência Computacional
Prof. Luís Alexandre



Índice

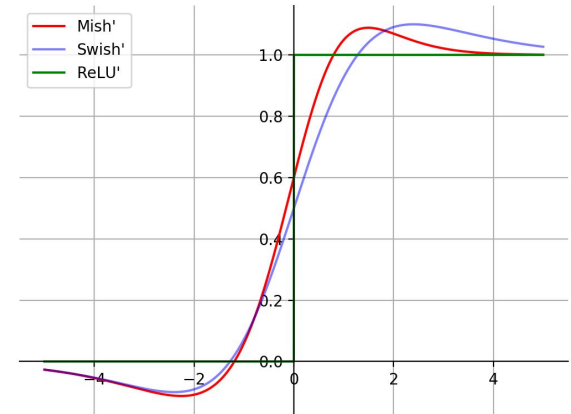
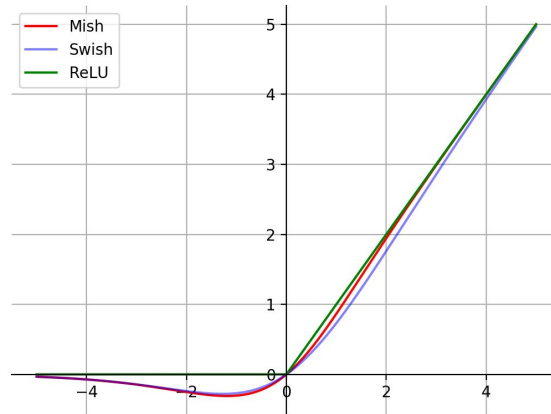
- Introdução;
- Função de Ativação Mish;
- SWA: Stochastic Weight Averaging;
- Testes e Resultados;
- Referências.

Introdução



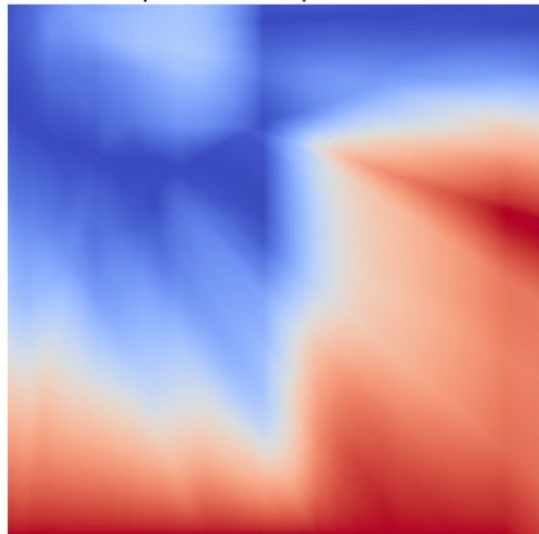
Função de Ativação Mish

- Não saturada;
- Suave e contínua;
- Diferenciável;
- Ativação para inputs negativos;
- Regularização intrínseca.

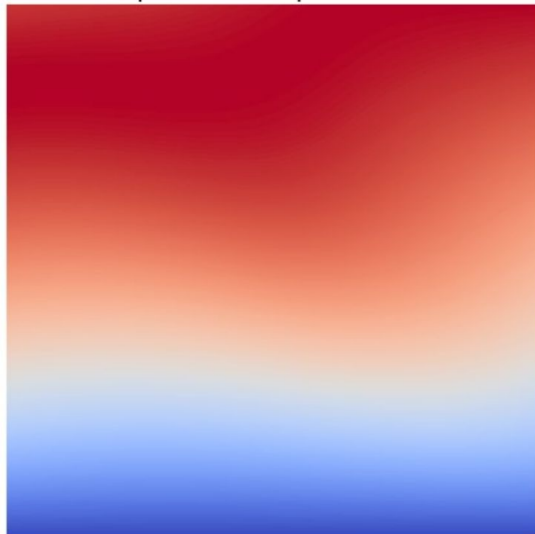


Função de Ativação Mish

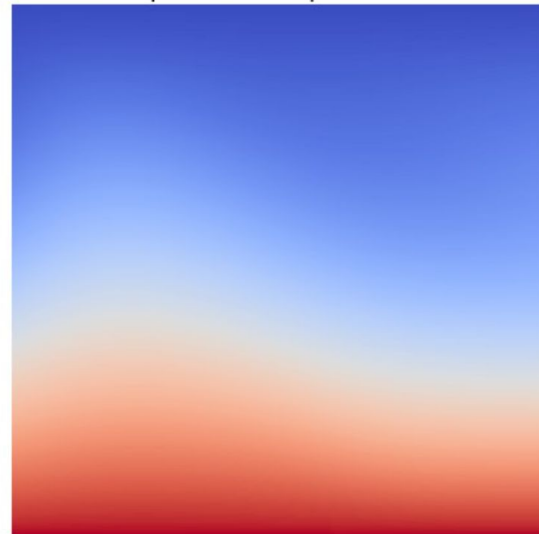
Output Landscape for ReLU



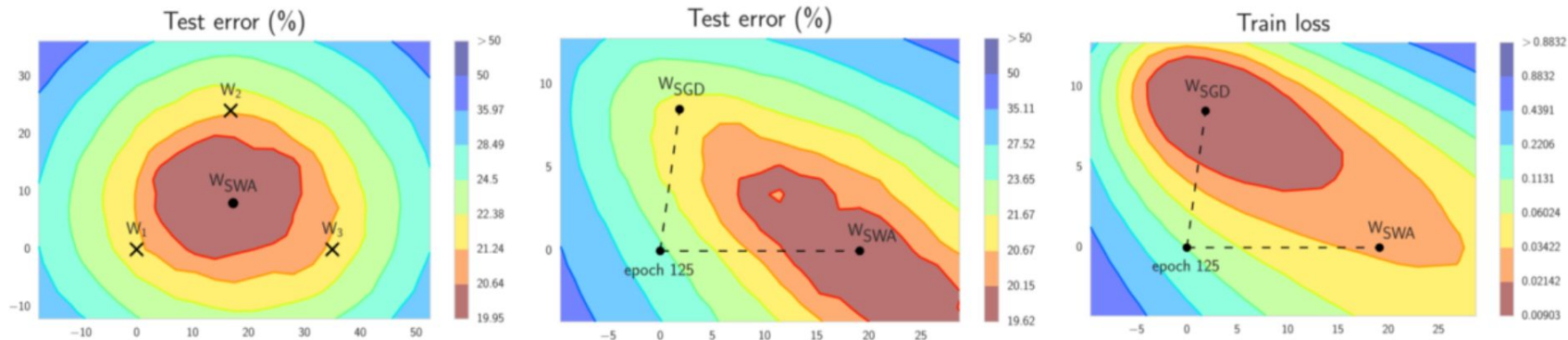
Output Landscape for Swish



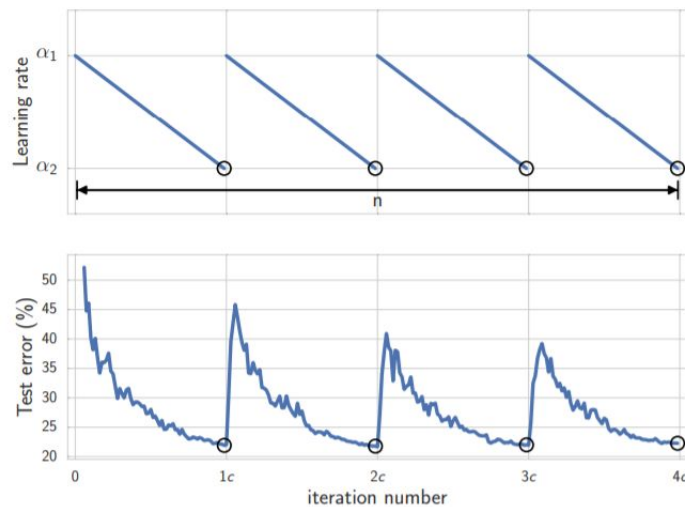
Output Landscape for Mish



SWA: Stochastic Weight Averaging



SWA: Stochastic Weight Averaging





SWA: Stochastic Weight Averaging

Algorithm 1 Algoritmo SWA

Require: pesos \hat{w} , valores da taxa de aprendizagem α_1 e α_2 , tamanho do ciclo c (para uma taxa de aprendizagem constante $c = 1$), número de iterações n

Ensure: w_{SWA}

$w \leftarrow \hat{w}$

$w_{SWA} \leftarrow w$

for $i \leftarrow 1, 2, \dots, n$ **do**

$\alpha \leftarrow \alpha(i)$

$w \leftarrow w - \alpha \nabla \mathcal{L}_I(w)$

if $\text{mod}(i, c) = 0$ **then**

$n_{\text{models}} \leftarrow i/c$

$w_{SWA} \leftarrow \frac{w_{SWA} \cdot n_{\text{models}} + w}{n_{\text{models}} + 1}$

end if

end for



Testes e Resultados (Mish: Replicação)

Modelo (Função de Ativação)	Resultado do <i>paper</i> (%)	Época Final - Replicado (%)	Melhor Época (%)	Tempo de Treino (s)
CIFAR-10				
Inception-V3 (Mish)	91.19	91.71 \pm 0.02	91.89 \pm 0.14	18991.24 \pm 15.29
DenseNet-121 (Mish)	91.27	90.72 \pm 0.27	91.09 \pm 0.10	9226.21 \pm 46.01
DenseNet-121 (Swish)	90.92	90.68 \pm 0.28	91.10 \pm 0.00	9604.06 \pm 355.49
DenseNet-121 (ReLU)	91.09	90.8 \pm 0.09	91.19 \pm 0.32	9590.13 \pm 51.02
CIFAR-100				
ResNext-50 (Mish)	67.58	68.19 \pm 0.15	68.86 \pm 0.19	12308.98 \pm 6.71
DenseNet-121 (Mish)	66.31	65.94 \pm 0.56	66.95 \pm 0.26	9246.68 \pm 22.69



Testes e Resultados (SWA: Replicação)

Modelo (Função de Ativação)	Resultado do <i>paper</i> (%)	Época Final - Replicado (%)	Melhor Época (%)	Tempo Médio de Treino (s)
CIFAR-10				
VGG-16 (ReLU)	93.59 \pm 0.16	93.53 \pm 0.08	93.57 \pm 0.08	1643.24 \pm 7.73
Resnet-164 (ReLU)	95.56 \pm 0.11	95.62 \pm 0.05	95.71 \pm 0.02	6968.53 \pm 117.24
CIFAR-100				
VGG-16 (ReLU)	73.91 \pm 0.12	73.67 \pm 0.25	73.67 \pm 0.25	1642.23 \pm 5.53
Resnet-164 (ReLU)	79.77 \pm 0.17	79.53 \pm 0.26	79.56 \pm 0.29	6830.03 \pm 2.78



Testes e Resultados (Implementação Híbrida)

Função de Ativação	α SGD	α SWA	Época Final (%)	Melhor Época (%)	Tempo Médio de Treino (s)
ReLU	0.05	0.01	93.43 \pm 0.02	93.46 \pm 0.002	1662.04 \pm 11.28
Mish	0.05	0.01	93.31 \pm 0.18	93.35 \pm 0.14	1662.24 \pm 4.68
ReLU	0.001	0.001	89.02 \pm 0.09	89.02 \pm 0.08	1654.35 \pm 5.34
Mish	0.001	0.001	88.59 \pm 0.07	88.61 \pm 0.05	1670.85 \pm 1.80
ReLU	0.01	0.01	92.45 \pm 0.16	92.50 \pm 0.14	1663.10 \pm 2.99
Mish	0.01	0.01	92.56 \pm 0.19	92.64 \pm 0.21	1745.10 \pm 56.96



Referências

- (Mish) <https://arxiv.org/pdf/1908.08681.pdf>;
- (SWA) <https://arxiv.org/pdf/1803.05407v3.pdf>

Dúvidas?