Aula 6: o MNIST em NumPy

27/11/2026, on-line. O que não couber entra na aula seguinte.

1. Revisão das aulas 4 e 5

de memória: o Jacobiano, de trás para a frente, o δ, a derivada numérica
o fio do curso, da aula 1 à 10; hoje, a 6

2. O MNIST e a rede

70 mil dígitos, 60 mil de treino e 10 mil de teste; 28 × 28 pixels, de 0 a 1
vídeo: o dígito vira uma lista de 784 números; 0,9998 no 7; 0,511 contra 0,489
a rede 784, 256, 10: tanh na oculta, softmax na saída; 203.530 números
softmax: $p_k = e^{z_k} / \Sigma\, e^{z_j}$; positivos, somam 1; (2; 1; 0) dá (0,665; 0,245; 0,090)
vídeo: a softmax por dentro; menos 2, as mesmas; vezes 3, (0,950; 0,047; 0,002)
entropia cruzada: $L = -\ln p_t$; 0,105 com 0,9; 2,303 com 0,1
retropropagação: δ da saída = p − y, caixa-preta; $\tanh' = 1 - \tanh^2$; $1{,}5 \cdot 10^{-8}$
lotes de 64: 782 passos por época; η = 0,1; 20 épocas

3. Treino, validação e teste

a rede aprendeu ou decorou? 99,21% no treino não diz
50 mil ajustam, 10 mil escolhem, 10 mil medem uma vez
curvas: 99,21% no treino, 97,33% na validação, época 20
sobreajuste com mil exemplos: 100% no treino, 87,42% na validação, mínimo na 23
larguras 16, 64, 256: 94,10%, 96,73%, 97,33%
teste: 97,59%, 241 erros; semente 1, versão 227111746c19
a matriz de confusão: 15 setes como 2, 12 quatros como 9; o 9 acerta menos, 96,3%
os pesos de 16 neurônios: traços, como o 5 e o 15, e ruído
vídeo: o neurônio 5 diante de um 7, soma −2,83, e de um 5, +1,55

4. De onde vem

NIST, SD-3 do censo e SD-1 do ensino médio; LeNet-5, 0,95%; MLP de 1998, 4,7%
Prêmio Turing de 2018: Bengio, Hinton e LeCun

5. O código

mlp.py só pela saída: é a referência da entrega
a entrega: o esqueleto com softmax, para_a_frente e retropropagacao escritas

6. Fechamento

PyTorch, Learn the Basics; Karpathy, micrograd; Kingma e Ba, o Adam
a seguir: aula 7, PyTorch; aula 8, os dados do CERN