Runge Function & Derivative — Joint Neural ApproximationTarget

$$f(x) = \frac{1}{1+25x^2}, \quad f'(x) = -\frac{50x}{(1+25x^2)^2}, \quad x \in [-1,1]$$

- Method

  Model 1-hidden-layer MLP (width H=96), smooth tanh-like activation ϕ with closed-form ϕ', ϕ'' ; linear output.

  Prediction/derivative

  $$\hat{f}(x) = b_2 + \sum_j W_j^{[2]} \phi(W_j^{[1]} x + b_j^{[1]}),$$
  $$\hat{f}'(x) = \sum_j W_j^{[2]} \phi'(W_j^{[1]} x + b_j^{[1]}) W_j^{[1]}$$

  Data 256 train points (uniform grid); 256 validation points (midpoints).

  Loss $\mathcal{L} = \mathrm{MSE}_f + \mu\,\mathrm{MSE}_{f'}$ with $\mu = 3$.

  Training: full-batch GD, lr =0.01 with step decays; 3000–3200 epochs.

- Results

  Loss curves (val ≈ train throughout)

  At representative epochs:

Epoch	Train tot	Val tot	Train MSE_f	Val MSE_f	Train $\mathrm{MSE}_{f'}$	Val $\mathrm{MSE}_{f'}$
1000	4.100	4.056	0.1960	0.1918	1.3014	1.2882
3000	**0.1046**	**0.1048**	**1.84e-4**	**1.84e-4**	**3.48e-2**	**3.49e-2**

  What the plots show (qualitative)

  $\hat{f}$ overlaps f on $[-1,1]$(errors < 10^{-3} on average).

  $\widehat{f}'$captures the odd S-shape with small bias where curvature changes fastest.

  Error to report (validation @ epoch 3000)

  Function MSE: 1.84×10^{-4}

  Derivative MSE: 3.49×10^{-2}

- Discussion

  Smooth ϕ + using ϕ', ϕ''in backprop lets the derivative term shape the hidden layer (avoids “constant derivative” failure of pure ReLU).

  Train/val match closely → good generalization.

  If needed to improve f' increase width to H=128, extend epochs with one more lr decay, or raise μ (e.g., $\mu = 5$); Chebyshev-biased sampling can also reduce edge errors.