

Exercises

§13.5,13.7: Training Neural Networks

Python: There is an optional starter script that supports Python-based problem(s) on this sheet. Use it as scaffolding: read the comments, complete the lines marked **TODO**, and compare the output with the mathematics requested in the problem(s).

Problem 1 Using the gradients from Lesson 1.33 and learning rate $\alpha = 0.01$, compute the updated W , $\mathbf{b}$, $\mathbf{v}$, and c . Then compute the new prediction and loss for $\mathbf{x} = (2, 1)^T$ and $y = 1.75$.

Problem 2 Repeat Problem 1 with $\alpha = 0.1$. Compute the new loss. Explain why the result does not contradict the fact that $-\nabla L$ is a descent direction.

Problem 3 A training set contains $P = 10,000$ observations and uses mini-batches of size $b = 128$. How many parameter updates occur in one epoch if the final partial batch is retained? How many updates occur in 20 epochs?

Problem 4 Explain why initializing every hidden unit with exactly the same incoming weight vector and bias prevents the units from learning distinct features under ordinary gradient-based training. Why does random initialization help?

Problem 5 Suppose a layer has $n_{\text{in}} = 100$ inputs. Compute the standard deviation corresponding to (a) variance $1/n_{\text{in}}$, and (b) variance $2/n_{\text{in}}$. Which scale is associated approximately with Xavier-style initialization and which with He initialization?

Problem 6 A network uses the regularized objective

$$g_{\text{reg}}(W) = g_{\text{data}}(W) + \frac{\lambda}{2} \|W\|_F^2.$$

Derive the regularized gradient and the gradient-descent update. Show the multiplicative shrinkage factor on the old weight matrix.

Problem 7 Training and validation losses are:

Epoch	Training Loss	Validation Loss
1	0.82	0.91
5	0.48	0.55
10	0.29	0.37
15	0.19	0.34
20	0.12	0.38
25	0.08	0.45

Which epoch has the best

validation loss? What evidence of overfitting appears? If early stopping stores the best checkpoint, which parameter state should be restored?

Problem 8 Implement the small $y = |x|$ ReLU training example from the lesson. Use the fixed seed and learning rate shown. Record the loss at epochs 0, 10, 50, 100, 500, 999. Then change the learning rate and describe how the training trajectory changes.