Numerical Problems in Deep Learning (All Units)

UNIT–I Numerical Problems

1. A perceptron receives inputs $(x_1, x_2) = (2, -1)$ with weights $(w_1, w_2) = (0.5, -0.25)$ and bias $b = 0.2$. Compute the perceptron output using the step activation.

2. For a dataset containing points belonging to two classes, a perceptron misclassifies $(x_1, x_2) = (1, -2)$ with desired output $y = 1$. If learning rate $\eta = 0.1$ and current weights are $(0.4, -0.3)$, compute the updated weights.

3. A sigmoid neuron computes

$$z = w_1x_1 + w_2x_2 + b$$

   Given $x_1 = 1$, $x_2 = 3$, $w_1 = 0.2$, $w_2 = -0.1$, $b = 0.5$, compute z and $\sigma(z)$.

4. For a loss function

$$L = \frac{1}{2}(t - y)^2,$$

   where $y = wx$, $t = 4$, $x = 2$, and $w = 1.5$, compute

$$\frac{\partial L}{\partial w}.$$

5. Given learning rate $\eta = 0.01$, update the weight using gradient descent from Q4.

UNIT–II Numerical Problems

[resume]

1. A network has output $y = \sigma(wx)$ where σ is sigmoid. Given $x = 0.5$, $w = 0.8$, and target $t = 1$, compute the gradient

$$\frac{\partial L}{\partial w}$$

   for squared error.

2. In Momentum-based GD, compute the new velocity and weight:

$$v_t = \beta v_{t-1} + (1 - \beta)\nabla L, \quad w_t = w_{t-1} - \eta v_t$$

   Given $\beta = 0.9$, $v_{t-1} = 0.4$, $\nabla L = 0.2$, $\eta = 0.01$, $w_{t-1} = 2$.

3. In RMSProp, compute the updated running average:

$$E[g^2]_t = 0.9E[g^2]_{t-1} + 0.1g_t^2$$

   Given $E[g^2]_{t-1} = 0.5$, $g_t = 0.3$.

4. Compute PCA: Given covariance matrix

$$C = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix},$$

   compute eigenvalues.

5. Perform SVD on

$$A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}.$$

   Find $A^T A$.

UNIT–III Numerical Problems

1. A denoising autoencoder input is $x = 0.8$ and noise $\epsilon = -0.2$. Compute the corrupted input $\tilde{x}$.

2. In L2 regularization, compute the regularized loss:

$$L_{reg} = L + \lambda \|w\|^2,$$

   where $L = 0.4$, $\lambda = 0.01$, $\|w\|^2 = 25$.

3. A dropout layer drops 40

4. Batch Normalization transforms

$$\hat{x} = \frac{x - \mu}{\sigma}, \quad y = \gamma \hat{x} + \beta$$

   Given $x = 10$, $\mu = 8$, $\sigma = 2$, $\gamma = 1.5$, $\beta = 0.5$, compute y.

5. A sparse autoencoder has reconstruction error 0.12 and sparsity penalty 0.03. Compute total cost.

UNIT–IV Numerical Problems

1. Perform 1D convolution: Input: $[1, 2, 3, 4]$ Filter: $[1, -1]$ Stride $= 1$. Compute output.

2. Given a 2×2 pooling window with max pooling on

$$\begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$$

   compute pooled value.

3. A CNN layer uses 32 filters of size $3 \times 3 \times 16$. Compute total trainable parameters (including bias).

4. In ResNet, a residual block has input $x = 5$ and $F(x) = -2$. Compute block output $y = F(x) + x$.

5. A feature map has dimension $28 \times 28 \times 64$. After applying 2×2 stride-2 max pooling, compute output size.

UNIT–V Numerical Problems

1. An RNN computes
$$h_t = \tanh(Wh_{t-1} + Ux_t)$$
Given $h_{t-1} = 0.5$, $x_t = 1$, $W = 0.2$, $U = 0.4$, compute h_t.

2. In BPTT, total gradient is
$$\frac{\partial L}{\partial w} = \sum_{t=1}^{3} \frac{\partial L_t}{\partial w}$$
If 0.3, 0.5, 0.2 are gradients for $t = 1, 2, 3$, compute the total.

3. In LSTM, forget gate is
$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$$
Given $x_t = 1$, $h_{t-1} = 2$, $W_f = 0.5$, $U_f = -0.1$, $b_f = 0$, compute f_t.

4. In attention mechanism, compute attention score
$$e = qk$$
Given $q = 0.6$, $k = 0.4$.

5. Compute softmax values for
$$z = [2, 1, 0]$$
to three decimal places.

Attention Scores based Problems

1. Given a query vector
$$q = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad k_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \quad k_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$
compute the unnormalized attention scores
$$e_1 = q^T k_1, \quad e_2 = q^T k_2.$$

2. Using the scores from Question 1, compute the attention weights
$$\alpha_i = \frac{\exp(e_i)}{\exp(e_1) + \exp(e_2)}, \quad i = 1, 2.$$
Express α_1 and α_2 numerically (you may keep e-based values or approximate up to 3 decimal places).

3. Let the value vectors be

$$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

Using the attention weights α_1, α_2 from Question 2, compute the context vector

$$c = \alpha_1 v_1 + \alpha_2 v_2.$$

4. Consider scaled dot-product attention. Given

$$q = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad k = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad d_k = 2,$$

compute the scaled attention score

$$e = \frac{q^T k}{\sqrt{d_k}}.$$

5. You are given three keys and one query:

$$q = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad k_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad k_2 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \quad k_3 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

   (a) Compute the unnormalized scores $e_i = q^T k_i$ for $i = 1, 2, 3$.

   (b) Compute the softmax attention weights

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{3} \exp(e_j)}, \quad i = 1, 2, 3.$$

6. Given attention logits (unnormalized scores)

$$e = [2, 1, 0],$$

compute the softmax attention weights

$$\alpha_i = \frac{\exp(e_i)}{\exp(2) + \exp(1) + \exp(0)}, \quad i = 1, 2, 3.$$

Provide the values up to 3 decimal places.

7. Masked attention: suppose we have scores

$$e = [3, -1, 0.5]$$

and we apply a mask that *disallows* the second position by assigning it $-\infty$. Conceptually, this is implemented as:

$$e' = [3, -\infty, 0.5].$$

   (a) Write the softmax expression for $\alpha_1, \alpha_2, \alpha_3$ using e'.

   (b) Explain why $\alpha_2 = 0$ and compute approximate values for α_1 and α_3 (up to 3 decimal places).

8. Multi-head style small example: Let the query and key matrices be

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad K = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Compute the score matrix

$$S = QK^T.$$

Write out S explicitly.

9. For a single query q and two key-value pairs (k_1, v_1) and (k_2, v_2), attention is defined as:

$$\text{Attention}(q, K, V) = \alpha_1 v_1 + \alpha_2 v_2,$$

where

$$\alpha_i = \frac{\exp(q^T k_i)}{\exp(q^T k_1) + \exp(q^T k_2)}.$$

Given:

$$q = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \; k_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \; k_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \; v_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \; v_2 = \begin{bmatrix} 0 \\ 3 \end{bmatrix},$$

compute the context vector $\text{Attention}(q, K, V)$ numerically.

10. Scaled dot-product attention in matrix form: Let

$$Q = \begin{bmatrix} 1 & 2 \end{bmatrix}, \quad K = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad V = \begin{bmatrix} 2 & 1 \\ 3 & 0 \end{bmatrix}, \quad d_k = 2.$$

(a) Compute the score vector

$$s = \frac{QK^T}{\sqrt{d_k}}.$$

(b) Apply softmax to s to get attention weights α.

(c) Compute the final output

$$\text{output} = \alpha V.$$