1 Neural Networks

1.1 Basic MLP

(a)

From Layer 0 to Layer 1: (2 inputs + 1 bias) × 3 neurons = 9 weights.

From Layer 1 to Layer 2: (3 neurons + 1 bias) × 3 neurons = 12 weights.

From Layer 2 to Layer 3: (3 neurons + 1 bias) × 1 neuron = 4 weights.

Total weights $= 9 + 12 + 4 = 25$

(b)

Layer 1:

$$\vec{z_1} = W_0\vec{x} = \begin{bmatrix} 0.2 & 0.5 & -1.0 \\ -0.7 & -0.7 & -0.3 \\ -0.1 & 0.9 & 0.8 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -0.3 \\ -1.7 \\ 1.6 \end{bmatrix} \implies \vec{h_1} = \begin{bmatrix} g(\vec{z_1}) \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1.6 \\ 1 \end{bmatrix}$$

Layer 2:

$$\vec{z_2} = W_1\vec{h_1} = \begin{bmatrix} -0.6 & -0.4 & 0.2 & -0.9 \\ -0.4 & -0.3 & 0.4 & 0.7 \\ -0.5 & 0.3 & -0.2 & -0.1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1.6 \\ 1 \end{bmatrix} = \begin{bmatrix} -0.58 \\ 1.34 \\ -0.42 \end{bmatrix} \implies \vec{h_2} = \begin{bmatrix} g(\vec{z_2}) \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1.34 \\ 0 \\ 1 \end{bmatrix}$$

Layer 3 (Output Layer):

$$\hat{y} = W_2\vec{h_2} = \begin{bmatrix} 1 & 0.9 & 0.6 & 0.9 \end{bmatrix} \begin{bmatrix} 0 \\ 1.34 \\ 0 \\ 1 \end{bmatrix} = 2.106$$

(c)

$$\frac{\partial L}{\partial W_2} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial W_2} \tag{1.1}$$

$$= (\hat{y} - y)\vec{h_2} \tag{1.2}$$

This is a vector of dimension 4.

(d)

$$
\begin{aligned}
\frac{\partial L}{\partial \vec{w}_{0,1,\cdot}^T} &= \begin{bmatrix}\vec{x} & \vec{0} & \vec{0}\end{bmatrix} \frac{\partial L}{\partial \vec{z_1}} && (2.1)\\
&= \begin{bmatrix}\vec{x} & \vec{0} & \vec{0}\end{bmatrix} \begin{bmatrix} g'(z_{1,1}) & 0 & 0 & 0\\ 0 & g'(z_{1,2}) & 0 & 0\\ 0 & 0 & g'(z_{1,3}) & 0\end{bmatrix} \frac{\partial L}{\partial \vec{h_1}} && (2.2)\\
&= \begin{bmatrix}\vec{x} & \vec{0} & \vec{0}\end{bmatrix} \begin{bmatrix} g'(z_{1,1}) & 0 & 0 & 0\\ 0 & g'(z_{1,2}) & 0 & 0\\ 0 & 0 & g'(z_{1,3}) & 0\end{bmatrix} W_1^T \frac{\partial L}{\partial \vec{z_2}} && (2.3)\\
&= \begin{bmatrix}\vec{x} & \vec{0} & \vec{0}\end{bmatrix} \begin{bmatrix} g'(z_{1,1}) & 0 & 0 & 0\\ 0 & g'(z_{1,2}) & 0 & 0\\ 0 & 0 & g'(z_{1,3}) & 0\end{bmatrix} W_1^T \begin{bmatrix} g'(z_{2,1}) & 0 & 0 & 0\\ 0 & g'(z_{2,2}) & 0 & 0\\ 0 & 0 & g'(z_{2,3}) & 0\end{bmatrix} W_2^T \frac{\partial L}{\partial \hat{y}} && (2.4)\\
&= \begin{bmatrix}\vec{x} & \vec{0} & \vec{0}\end{bmatrix} \cdot \prod_{l=1}^{2} \left[\begin{bmatrix}\text{diag}(g'(z_{l,1}), g'(z_{l,2}), g'(z_{l,3})) & \vec{0}\end{bmatrix} \cdot W_l^T\right] \cdot (\hat{y} - y) && (2.5)
\end{aligned}
$$

$$
\frac{\partial L}{\partial \vec{w}_{0,2,\cdot}^T} = \begin{bmatrix}\vec{0} & \vec{x} & \vec{0}\end{bmatrix} \cdot \prod_{l=1}^{2} \left[\begin{bmatrix}\text{diag}(g'(z_{l,1}), g'(z_{l,2}), g'(z_{l,3})) & \vec{0}\end{bmatrix} \cdot W_l^T\right] \cdot (\hat{y} - y) \tag{3}
$$

$$
\frac{\partial L}{\partial \vec{w}_{0,3,\cdot}^T} = \begin{bmatrix}\vec{0} & \vec{0} & \vec{x}\end{bmatrix} \cdot \prod_{l=1}^{2} \left[\begin{bmatrix}\text{diag}(g'(z_{l,1}), g'(z_{l,2}), g'(z_{l,3})) & \vec{0}\end{bmatrix} \cdot W_l^T\right] \cdot (\hat{y} - y) \tag{4}
$$

They are all vectors of dimension 3.

1.2 Revised MLP-1: MLP with specific forward directions

(a)

Layer 1:

$$\vec{z_1} = W_0\vec{x} = \begin{bmatrix} 0.7 & \star & \star \\ 0.3 & 0.5 & \star \\ 0.3 & 0.4 & -0.5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.7 \\ 0.8 \\ 0.2 \end{bmatrix} \implies \vec{h_1} = g(\vec{z_1}) = \begin{bmatrix} 0.7 \\ 0.8 \\ 0.2 \end{bmatrix}$$

Layer 2 (Output Layer):

$$\hat{\vec{y}} = W_1\vec{h_1} = \begin{bmatrix} 0.4 & \star & \star \\ 0.1 & -0.9 & \star \\ -0.9 & 0.8 & 0.2 \end{bmatrix} \begin{bmatrix} 0.7 \\ 0.8 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 0.28 \\ -0.65 \\ 0.05 \end{bmatrix}$$

(b)

$$\begin{aligned}
\frac{\partial L}{\partial \vec{w}_{0,1,\cdot}^T} &= \begin{bmatrix} \vec{x} & \vec{0} & \vec{0} \end{bmatrix} \frac{\partial L}{\partial \vec{z_1}} && (5.1) \\
&= \begin{bmatrix} \vec{x} & \vec{0} & \vec{0} \end{bmatrix} \begin{bmatrix} g'(z_{1,1}) & 0 & 0 \\ 0 & g'(z_{1,2}) & 0 \\ 0 & 0 & g'(z_{1,3}) \end{bmatrix} \frac{\partial L}{\partial \vec{h_1}} && (5.2) \\
&= \begin{bmatrix} \vec{x} & \vec{0} & \vec{0} \end{bmatrix} \begin{bmatrix} g'(z_{1,1}) & 0 & 0 \\ 0 & g'(z_{1,2}) & 0 \\ 0 & 0 & g'(z_{1,3}) \end{bmatrix} W_1^T \frac{\partial L}{\partial \hat{\vec{y}}} && (5.3) \\
&= \begin{bmatrix} \vec{x} & \vec{0} & \vec{0} \end{bmatrix} \cdot \text{diag}(g'(z_{1,1}), g'(z_{1,2}), g'(z_{1,3})) \cdot W_1^T \cdot 2(\hat{\vec{y}} - \vec{y}) && (5.4)
\end{aligned}$$

$$\frac{\partial L}{\partial \vec{w}_{0,2,\cdot}^T} = \begin{bmatrix} \vec{0} & \vec{x} & \vec{0} \end{bmatrix} \cdot \text{diag}(g'(z_{1,1}), g'(z_{1,2}), g'(z_{1,3})) \cdot W_1^T \cdot 2(\hat{\vec{y}} - \vec{y}) \tag{6}$$

$$\frac{\partial L}{\partial \vec{w}_{0,3,\cdot}^T} = \begin{bmatrix} \vec{0} & \vec{0} & \vec{x} \end{bmatrix} \cdot \text{diag}(g'(z_{1,1}), g'(z_{1,2}), g'(z_{1,3})) \cdot W_1^T \cdot 2(\hat{\vec{y}} - \vec{y}) \tag{7}$$

They are all vectors of dimension 3.

1.3 Revised MLP-2: MLP with shared weights

(a)

Layer 1:

$$\vec{z_1} = W\vec{x} = \begin{bmatrix} -0.5 & 0.7 & 0.2 & -0.8 \\ 1.0 & 1.0 & 0.1 & 0.4 \\ -0.2 & -0.4 & 0.2 & 0.4 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -0.4 \\ 2.5 \\ 0 \end{bmatrix} \implies \vec{h_1} = \begin{bmatrix} g(\vec{z_1}) \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2.5 \\ 0 \\ 1 \end{bmatrix}$$

Layer 2 (Output Layer):

$$\hat{\vec{y}} = W\vec{h_1} = \begin{bmatrix} -0.5 & 0.7 & 0.2 & -0.8 \\ 1.0 & 1.0 & 0.1 & 0.4 \\ -0.2 & -0.4 & 0.2 & 0.4 \end{bmatrix} \begin{bmatrix} 0 \\ 2.5 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.95 \\ 2.9 \\ -0.6 \end{bmatrix}$$

(b)

$$\frac{\partial L}{\partial \vec{w}_{0,1,\cdot}^T} = \begin{bmatrix} \vec{x} & \vec{0} & \vec{0} \end{bmatrix} \frac{\partial L}{\partial \vec{z_1}} \tag{8.1}$$

$$= \begin{bmatrix} \vec{x} & \vec{0} & \vec{0} \end{bmatrix} \begin{bmatrix} g'(z_{1,1}) & 0 & 0 & 0 \\ 0 & g'(z_{1,2}) & 0 & 0 \\ 0 & 0 & g'(z_{1,3}) & 0 \end{bmatrix} \frac{\partial L}{\partial \vec{h_1}} \tag{8.2}$$

$$= \begin{bmatrix} \vec{x} & \vec{0} & \vec{0} \end{bmatrix} \begin{bmatrix} g'(z_{1,1}) & 0 & 0 & 0 \\ 0 & g'(z_{1,2}) & 0 & 0 \\ 0 & 0 & g'(z_{1,3}) & 0 \end{bmatrix} W^T \frac{\partial L}{\partial \hat{\vec{y}}} \tag{8.3}$$

$$= \begin{bmatrix} \vec{x} & \vec{0} & \vec{0} \end{bmatrix} \cdot \begin{bmatrix} \text{diag}(g'(z_{1,1}), g'(z_{1,2}), g'(z_{1,3})) & \vec{0} \end{bmatrix} \cdot W^T \cdot 2(\hat{\vec{y}} - \vec{y}) \tag{8.4}$$

$$\frac{\partial L}{\partial \vec{w}_{0,2,\cdot}^T} = \begin{bmatrix} \vec{0} & \vec{x} & \vec{0} \end{bmatrix} \cdot \begin{bmatrix} \text{diag}(g'(z_{1,1}), g'(z_{1,2}), g'(z_{1,3})) & \vec{0} \end{bmatrix} \cdot W^T \cdot 2(\hat{\vec{y}} - \vec{y}) \tag{9}$$

$$\frac{\partial L}{\partial \vec{w}_{0,3,\cdot}^T} = \begin{bmatrix} \vec{0} & \vec{0} & \vec{x} \end{bmatrix} \cdot \begin{bmatrix} \text{diag}(g'(z_{1,1}), g'(z_{1,2}), g'(z_{1,3})) & \vec{0} \end{bmatrix} \cdot W^T \cdot 2(\hat{\vec{y}} - \vec{y}) \tag{10}$$

They are all vectors of dimension 4.

By transposing each $\frac{\partial L}{\partial \vec{w}_{0,j,\cdot}^T}$ for $j = 1, 2, 3$

$$\frac{\partial L}{\partial W} = \begin{bmatrix} \frac{\partial L}{\partial \vec{w}_{0,1,\cdot}^T}^T \\ \frac{\partial L}{\partial \vec{w}_{0,2,\cdot}^T}^T \\ \frac{\partial L}{\partial \vec{w}_{0,3,\cdot}^T}^T \end{bmatrix}$$

The result is a 3x4 matrix.

2 SVM

(a)

Square loss of $(x_1, y_1) = (-1 - (\begin{bmatrix}0 & 0.5\end{bmatrix}\begin{bmatrix}2\\0\end{bmatrix} - 0.5))^2 = 0.25$

Square loss of $(x_2, y_2) = (1 - (\begin{bmatrix}0 & 0.5\end{bmatrix}\begin{bmatrix}-2\\0\end{bmatrix} - 0.5))^2 = 2.25$

Average square loss $= \frac{0.25+2.25}{2} = 1.25$

Hinge loss of $(x_1, y_1) = \max(0, 1 - (-1)(\begin{bmatrix}0 & 0.5\end{bmatrix}\begin{bmatrix}2\\0\end{bmatrix} - 0.5)) = 0.5$

Hinge loss of $(x_2, y_2) = \max(0, 1 - (1)(\begin{bmatrix}0 & 0.5\end{bmatrix}\begin{bmatrix}-2\\0\end{bmatrix} - 0.5)) = 1.5$

Average hinge loss $= \frac{0.5+1.5}{2} = 1$

(b)

Square loss of $((5,5), -1) = (-1 - (\begin{bmatrix}0 & 0.5\end{bmatrix}\begin{bmatrix}5\\5\end{bmatrix} - 0.5))^2 = 9$

Average square loss $= \frac{0.25+2.25+9}{3} = \frac{23}{6}$

Hinge loss of $((5,5), -1) = \max(0, 1 - (-1)(\begin{bmatrix}0 & 0.5\end{bmatrix}\begin{bmatrix}5\\5\end{bmatrix} - 0.5)) = 3$

Average hinge loss $= \frac{0.5+1.5+3}{3} = \frac{5}{3}$

(c)

$$\overrightarrow{w^*} = \begin{bmatrix}-\frac{1}{2}\\0\end{bmatrix}, b^* = 0$$

Because it will correctly classify the 3 points, while passing through the mid point of $(2,0)$ and $(-2,0)$, and perpendicular to the line between $(2,0)$ and $(-2,0)$ keeping maximal margin.

(d)

$$\mathcal{L}(\vec{\lambda}) = \sum_{i=1}^{3} \lambda_i - \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3} \lambda_i\lambda_j y_i y_j k(\vec{x_i}, \vec{x_j})$$

$$= \sum_{i=1}^{3} \lambda_i - \frac{1}{2}[\lambda_1^2 y_1^2 k(x_1, x_1) + \lambda_2^2 y_2^2 k(x_2, x_2) + \lambda_3^2 y_3^2 k(x_3, x_3) +$$
$$2\lambda_1\lambda_2 y_1 y_2 k(x_1, x_2) + 2\lambda_1\lambda_3 y_1 y_3 k(x_1, x_3) + 2\lambda_2\lambda_3 y_2 y_3 k(x_2, x_3)]$$

Substituting $y_1 = -1, y_2 = 1, y_3 = -1$

$$= \sum_{i=1}^{3} \lambda_i - \frac{1}{2}[\lambda_1^2 k(x_1, x_1) + \lambda_2^2 k(x_2, x_2) + \lambda_3^2 k(x_3, x_3) -$$
$$2\lambda_1\lambda_2 k(x_1, x_2) + 2\lambda_1\lambda_3 k(x_1, x_3) - 2\lambda_2\lambda_3 k(x_2, x_3)]$$

Substituting values of $k(x_i, x_j)$
$k(x_i, x_j) = \text{exp}(0) = 1$ for $i = j$
$k(x_1, x_2) = \text{exp}(-\frac{1}{2} \cdot 4^2) = \text{exp}(-8)$
$k(x_1, x_3) = \text{exp}(-\frac{1}{2} \cdot 34) = \text{exp}(-17)$
$k(x_2, x_3) = \text{exp}(-\frac{1}{2} \cdot 74) = \text{exp}(-37)$
Therefore

$$= \lambda_1 + \lambda_2 + \lambda_3 - \frac{1}{2}[\lambda_1^2 + \lambda_2^2 + \lambda_3^2 - 2\lambda_1\lambda_2\text{exp}(-8) + 2\lambda_1\lambda_3\text{exp}(-17) - 2\lambda_2\lambda_3\text{exp}(-37)]$$

For each λ_i, we want to maximize the objectives:

$$L_1(\lambda_1) = \lambda_1 - \frac{1}{2}[\lambda_1^2 - 2\lambda_1\lambda_2\text{exp}(-8) + 2\lambda_1\lambda_3\text{exp}(-17)]$$

$$L_2(\lambda_2) = \lambda_2 - \frac{1}{2}[\lambda_2^2 - 2\lambda_1\lambda_2\text{exp}(-8) - 2\lambda_2\lambda_3\text{exp}(-37)]$$

$$L_3(\lambda_3) = \lambda_2 - \frac{1}{2}[\lambda_3^2 + 2\lambda_1\lambda_3\text{exp}(-17) - 2\lambda_2\lambda_3\text{exp}(-37)]$$

Now we find the derivatives and double derivatives of L_i with respect to λ_i

$$\frac{\partial L_1}{\partial \lambda_1} = 1 - \lambda_1 + \lambda_2 \text{exp}(-8) - \lambda_3 \text{exp}(-17), \quad \frac{\partial^2 L_1}{\partial \lambda_1^2} = -1$$

$$\frac{\partial L_2}{\partial \lambda_2} = 1 - \lambda_2 + \lambda_1 \text{exp}(-8) + \lambda_3 \text{exp}(-37), \quad \frac{\partial^2 L_2}{\partial \lambda_2^2} = -1$$

$$\frac{\partial L_3}{\partial \lambda_3} = 1 - \lambda_3 - \lambda_1 \text{exp}(-17) + \lambda_2 \text{exp}(-37), \quad \frac{\partial^2 L_3}{\partial \lambda_3^2} = -1$$

Since $\frac{\partial^2 L_i}{\partial \lambda_i^2} < 0 \quad \forall i, L_i(\lambda_i)$ are all concave, solving $\frac{\partial L_i}{\partial \lambda_i} = 0$ for each i can give optimal dual variables $\vec{\lambda}$

$$\begin{bmatrix} 1 & -\text{exp}(-8) & \text{exp}(-17) \\ -\text{exp}(-8) & 1 & -\text{exp}(-37) \\ \text{exp}(-17) & -\text{exp}(-37) & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\vec{\lambda} \approx \begin{bmatrix} 1.00033553380146 \\ 1.00033557518695 \\ 0.999999958586732 \end{bmatrix}$$

(e)

$$g(\vec{x}_{test}) = \sum_{i=1}^{3} [\lambda_i^* y_i k(\vec{x_i}, \vec{x}_{test})] - \frac{1}{2} \max_{i:y=-1} \left(\sum_{j=1}^{3} \lambda_j^* y_j k(\vec{x_j}, \vec{x_i}) \right) - \frac{1}{2} \min_{i:y=1} \left(\sum_{j=1}^{3} \lambda_j^* y_j k(\vec{x_j}, \vec{x_i}) \right)$$

$$\approx -0.361262781607078$$

(f)

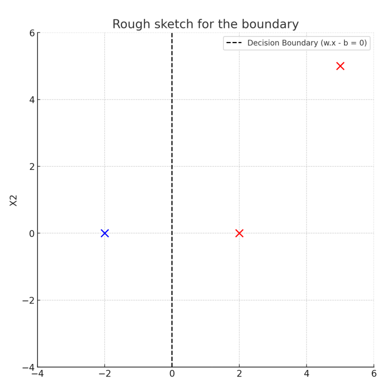

Figure 1: Rough sketch of decision boundary