Name:

Deep Learning
Homework 1

STN:

1)

Design a three layer neural network whose decision boundary is as shown in Figure 1. The gray region belongs to class 1 and other region belongs to class 0. Show your network structure, weights and nonlinear active
ation function.

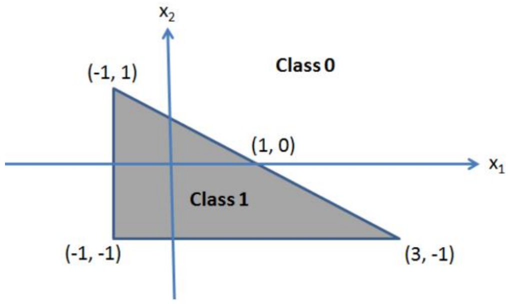

Figure 1:

2)

Consider the following function of $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5, x_6,)$:

$$f(\mathbf{x}) = \sigma\left(\log\left(5\left(\max\{x_1, x_2\} \cdot \frac{x_3}{x_4} - (x_5 + x_6)\right)\right) + \frac{1}{2}\right) \tag{9}$$

where σ is the sigmoid function

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{10}$$

Evaluate $f(\cdot)$ at $\hat{\mathbf{x}} = (5, -1, 6, 12, 7, -5)$. Then, compute the gradient $\nabla_{\mathbf{x}} f(\cdot)$ and evaluate it at the same point.

3)

Least-squares derivatives. Let $X_1, \ldots, X_N \in \mathbb{R}^p$ and $Y_1, \ldots, Y_N \in \mathbb{R}$. Define

$$X = \begin{bmatrix} X_1^\intercal \\ \vdots \\ X_N^\intercal \end{bmatrix} \in \mathbb{R}^{N \times p}, \qquad Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix} \in \mathbb{R}^N.$$

Let

$$\ell_i(\theta) = \frac{1}{2}(X_i^\intercal \theta - Y_i)^2 \quad \text{for } i = 1, \ldots, N, \qquad \mathcal{L}(\theta) = \frac{1}{2}\|X\theta - Y\|^2.$$

Show (a) $\nabla_\theta \ell_i(\theta) = (X_i^\intercal \theta - Y_i)X_i$ and (b) $\nabla_\theta \mathcal{L}(\theta) = X^\intercal(X\theta - Y)$.

Hint. For part (a), start by computing $\frac{\partial}{\partial \theta_j}\ell_i(\theta)$. For part (b), use the fact that

$$Mv = \sum_{i=1}^{N} M_{:,i} v_i \in \mathbb{R}^p$$

for any $M \in \mathbb{R}^{p \times N}$, $v \in \mathbb{R}^N$, where $M_{:,i}$ is the ith column of M for $i = 1, \ldots, N$.