Yunfan Yang
CS 442
February 25, 2022

HW1

Question 1.

$\mathcal{L}(w) := -\sum_{i=1}^{n} y_i \log \sigma\left(w \cdot x_i\right) + (1 - y_i) \log \sigma\left(-w \cdot x_i\right)$
$\mathcal{L}(w) := -\sum_{i=1}^{n} y_i \log \sigma\left(w \cdot x_i\right) + \log \sigma\left(-w \cdot x_i\right) - y_i \log \sigma\left(-w \cdot x_i\right)$
$\nabla\mathcal{L}(w) := -\sum_{i=1}^{n} y_i \frac{\delta(\sigma(w \cdot x_i))}{\sigma(w \cdot x_i)} + \frac{\delta(\sigma(-w \cdot x_i))}{\sigma(-w \cdot x_i)} - y_i \frac{\delta(\sigma(-w \cdot x_i))}{\sigma(-w \cdot x_i)}$
$\nabla\mathcal{L}(w) := -\sum_{i=1}^{n} y_i \frac{(1-\sigma(w \cdot x_i))\sigma(w \cdot x_i)x_i}{\sigma(w \cdot x_i)} + \frac{(1-\sigma(-w \cdot x_i))\sigma(-w \cdot x_i)(-x_i)}{\sigma(-w \cdot x_i)} - y_i \frac{(1-\sigma(-w \cdot x_i))\sigma(-w \cdot x_i)(-x_i)}{\sigma(-w \cdot x_i)}$
$\nabla\mathcal{L}(w) := -\sum_{i=1}^{n} y_i \left(1 - \sigma\left(w \cdot x_i\right)\right) x_i + \left(1 - \sigma\left(-w \cdot x_i\right)\right)\left(-x_i\right) - y_i \left(1 - \sigma\left(-w \cdot x_i\right)\right)\left(-x_i\right)$
$\nabla\mathcal{L}(w) := -\sum_{i=1}^{n} y_i x_i - y_i x_i \sigma\left(w \cdot x_i\right) - x_i + x_i \sigma\left(-w \cdot x_i\right) + y_i x_i - y_i x_i \sigma\left(-w \cdot x_i\right)$
$\nabla\mathcal{L}(w) := -\sum_{i=1}^{n} y_i x_i - \left(y_i x_i \sigma\left(w \cdot x_i\right) + y_i x_i \sigma\left(-w \cdot x_i\right)\right) - x_i + x_i \sigma\left(-w \cdot x_i\right) + y_i x_i$
$\nabla\mathcal{L}(w) := -\sum_{i=1}^{n} -x_i + x_i \sigma\left(-w \cdot x_i\right) + y_i x_i$
$\nabla\mathcal{L}(w) := -\sum_{i=1}^{n} x_i y_i - x_i \sigma\left(w \cdot x_i\right)$

Question 2.

2.1

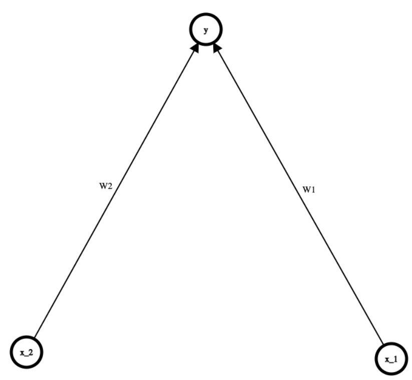

FIGURE 1. Two layer network.

$y = w_5(w_1 x_1 + w_2 x_2) + w_6(w_3 x_1 + w_4 x_2) + w_0$
$y = w_5 w_1 x_1 + w_5 w_2 x_2 + w_6 w_3 x_1 + w_6 w_4 x_2 + w_0$
$y = (w_5 w_1 + w_6 w_3) x_1 + (w_5 w_2 + w_6 w_4) x_2 + w_0$
In the new two-layer network, the weights are: $W1 = (w_5 w_1 + w_6 w_3), W1 = (w_5 w_2 + w_6 w_4)$

2.2

Since every three-layer network can be written into a two-layer network, i will use the model from 2.1 to prove
we get the following:
$W_1 + W_2 + W_0 \geq 0$
$-W_1 + W_2 + W_0 < 0$
$W_1 - W_2 + W_0 < 0$
$-W_1 - W_2 + W_0 \geq 0$
After some simplification, i found a contradiction, thus NO weight configuration can classify the XOR pattern without error:
$W_0 \geq 0$
$W_0 < 0$

2.3

$y_1 = sign(x_1 + x_2 - 0.5)$ *, this is identical to an OR function*
$y_2 = -1 * sign(x_1 + x_2 - 1.5)$ *, this is identical to a NAND function*
This Results an XOR gate if we set $w_5 = 0, w_6 = 0$

Question 3.

3.1

$h(x, a) = a$, this function achieves 0 classification error since $\Pr(Y = a \mid A = a) = 1$

3.2.1

error rate is $0.5 * p + 0.5 * (1 - p) = 0.5$

3.2.2

The above randomized classifier satisfies statistically parity because the coin is fair for each group, $Pr(\hat{Y} = 1) = 0.5$ for each group

3.3.1

suppose this deterministic classifier has the following traits:
$\Pr(\widehat{Y} = 1 \mid A = 0) = \alpha$
$\Pr(\widehat{Y} = 1 \mid A = 1) = 1 - \alpha$
Then,
$\mathbb{E}_\mu[h(X, A) \neq Y \mid A = 0] + \mathbb{E}_\mu[h(X, A) \neq Y \mid A = 1] = 1 - \Pr(\widehat{Y} = 1 \mid A = 1) + 1 - \Pr(\widehat{Y} = 0 \mid A = 0) = 1 + 1 - (\alpha + 1 - \alpha) = 1$

3.3.2

error $= p * \alpha + (1 - p) * (1 - \alpha) \geq (1 - p) * \alpha + (1 - p) * (1 - \alpha) = 1 - p$

Question 4.