CS 486/686 Introduction to Artificial Intelligence Spring 2015, Assignment 4 Solution

1 Question 1 (15 pts)

5 pts for each part:

$$\begin{aligned}\sigma(-a) &= \frac{1}{1+e^{a}}\\ &= \frac{e^{-a}}{e^{-a}+1}\\ &= \frac{1+e^{-a}-1}{e^{-a}+1}\\ &= \frac{1+e^{-a}}{1+e^{-a}} - \frac{1}{1+e^{-a}}\\ &= 1-\sigma(a)\end{aligned}$$

$$\begin{aligned}\sigma^{-1}(a) &= ln(a/(1-a))\\ a &= \sigma(ln(a/(1-a)))\\ &= \frac{1}{1+e^{-ln(a/(1-a))}}\\ &= \frac{1}{1+(1-a)/a}\\ &= \frac{1}{1/a}\\ &= a\end{aligned}$$

$$\begin{aligned}\frac{\partial\sigma}{\partial a} &= \frac{-1}{(1+e^{-a})^2}(-e^{-a})\\ &= \frac{1}{1+e^{-a}}\frac{e^{-a}}{1+e^{-a}}\\ &= \frac{1}{1+e^{-a}}\frac{1}{1+e^{a}}\\ &= \sigma(a)(1-\sigma(a))\end{aligned}$$

2 Question 2 (15 pts)

7 pts for showing the relation between $tanh(a)$ and $\sigma(a)$:

$$\begin{aligned}tanh(a) &= \frac{e^{a}-e^{-a}}{e^{a}+e^{-a}}\\ &= \frac{1-e^{-2a}}{1+e^{-2a}}\\ &= 2\sigma(2a)-1\end{aligned}$$

$$\sigma(a) = \frac{1}{2}tanh(\frac{a}{2}) + \frac{1}{2}$$

8 pts for showing the equivalent neural network and linear transformation of weights, now rewrite $y_i(x, W)$ in terms of $tanh()$:

$$\begin{aligned}y_i(x,W) &= \sigma\left(\sum_j W_{ji}^{(2)}\sigma\left(\sum_k W_{kj}^{(1)}x_k + W_{0j}^{(1)}\right) + W_{0i}^{(2)}\right)\\ &= \sigma\left(\sum_j W_{ji}^{(2)}\left[\frac{1}{2}tanh\left(\frac{1}{2}\sum_k W_{kj}^{(1)}x_k + \frac{1}{2}W_{0j}^{(1)}\right) + \frac{1}{2}\right] + W_{0i}^{(2)}\right)\\ &= \sigma\left(\sum_j \frac{1}{2}W_{ji}^{(2)}tanh\left(\sum_k \frac{1}{2}W_{kj}^{(1)}x_k + \frac{1}{2}W_{0j}^{(1)}\right) + \sum_j \frac{1}{2}W_{ji}^{(2)} + W_{0i}^{(2)}\right)\\ &= \sigma\left(\sum_j V_{ji}^{(2)}tanh\left(\sum_k V_{kj}^{(1)}x_k + V_{0j}^{(1)}\right) + V_{0i}^{(2)}\right)\end{aligned}$$

where V is obtained by a linear transformation of W:

$$V_{ji}^{(2)} = \frac{1}{2}W_{ji}^{(2)}$$

$$V_{0i}^{(2)} = \sum_j \frac{1}{2} W_{ji}^{(2)} + W_{0i}^{(2)}$$

$$V_{kj}^{(1)} = \frac{1}{2} W_{kj}^{(1)}$$

$$V_{0j}^{(1)} = \frac{1}{2} W_{0j}^{(1)}$$

3 Question 3 (30 pts)

- Is the dataset linearly separable? Explain briefly. (10 pts) Yes, the dataset is linearly separable as the perceptron finds a linear separator and the training accuracy converges to 100%.
- Train and test accuracy of the threshold perceptron. (10 pts) The accuracy is 100% for training and 90% for testing (your numbers might be slightly different).
- A printout of the final weights of the threshold perceptron. (10 pts) Here is a set of valid final weights: 99, 187, -270, -147, -81, -406, -125, 84, 153, -180, -128, 40, -339, -507, -353, -96, 13, -185, -112, -65, -187, -23, -160, -106, -234, -66, 87, -186, -231, -272, 4, -252, 112, -128, 63, 3, 231, 17, -128, 42, 158, 25, 724, 312, 235, -93, 389, -210, -329, -192, 174, 541, -142, 207, 264, 81, -114, -353, -262, -167, 205, 429, 46, 107, -33, you will get marks for weights in a similar range that makes sense.

4 Question 4

- Graph of train and test accuracy. (20 pts)
- Brief discussion of the results in the graph. (10 pts)
  - Training accuracy improves slightly with the number of hidden nodes, but it is not clear that this translates into better generalization, i.e. better test accuracy.
- Which algorithm performs best? Why? (10 pts)
  - Neural network works better because it produces a non-linear separator while the perceptron produces a linear separator.