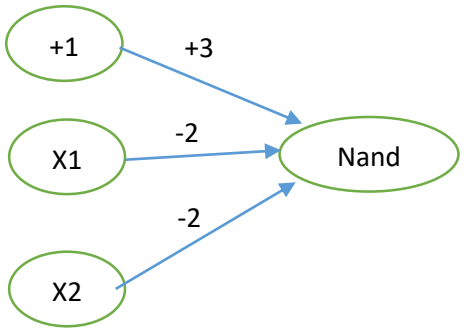
Yihan Zhou

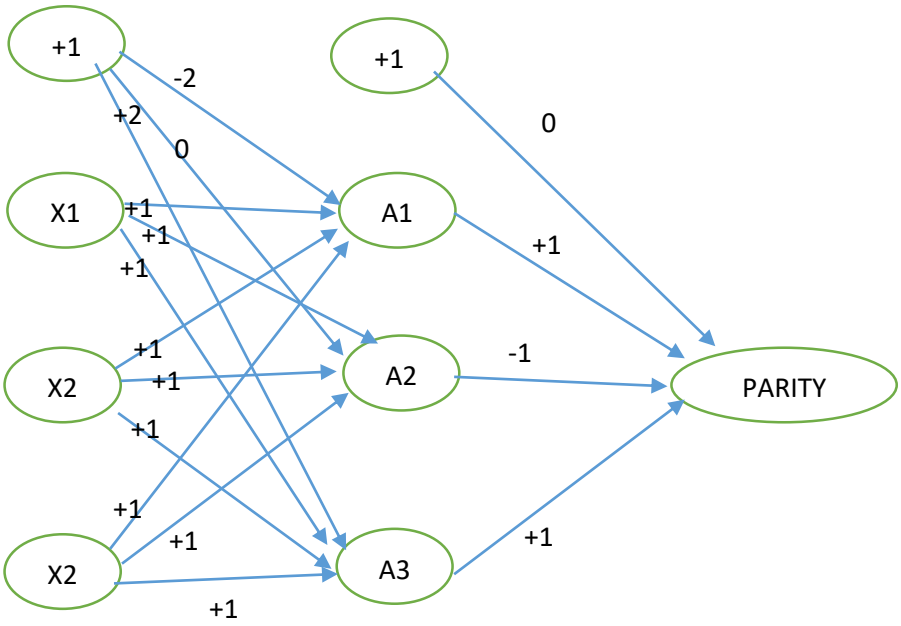
903053761

HW4

1.(a)



1.(b)



2.

1 round：

1st sample:

$o_h = g(0.1 * 1 + 0.1 * 0 + 0.1 * 1) = 0.54983399731$

$o_y = g(0.1 * o_h + 0.1 * 1) = 0.53866847996$

$\delta_y = (o_y - 1) = -0.01141871967$

$\delta_h = 0.1 * (\delta_y) * (1 - o_h) * o_h = -0.01141871967$

$\Delta\Theta^1_{0,h} = \Delta\Theta^1_{0,h} + 1 * \delta_h = -0.01141871967$

$\Delta\Theta^1_{x1,h} = \Delta\Theta^1_{x1,h} + x1 * \delta_h = -0.01141871967$

$\Delta\Theta^1_{x2,h} = \Delta\Theta^1_{x2,h} + x2 * \delta_h = 0$

$\Delta\Theta^2_{0,y} = \Delta\Theta^1_{0,y} + 1 * \delta_y = -0.46133152004$

$\Delta\Theta^2_{h,y} = \Delta\Theta^2_{h,y} + o_h * \delta_y = -0.25365575374$

2nd sample:

$o_h = g(0.1 * 1 + 0 + 0.1 * 1) = 0.54983399731$

$o_y = g(o_h * 1 + 0.1 * 1) = 0.53866847996$

$\delta_y = (o_y - 0) = 0.53866847996$

$\delta_h = 0.1 * (o_y) * (1 - o_h) * o_h = 0.01333293759$

$\Delta\Theta^1_{x2,h} = \Delta\Theta^1_{x2,h} + x2 * \delta_h = -0.01333293759$

$\Delta\Theta^1_{x1,h} = \Delta\Theta^1_{x1,h} + x1 * \delta_h = -0.01141871967$

$\Delta\Theta^1_{0,h} = \Delta\Theta^1_{0,h} + 1 * \delta_h = 0.00191421792$

$\Delta\Theta^2_{h,y} = \Delta\Theta^2_{h,y} + o_h * \delta_y = 0.04252248982$

$\Delta\Theta^2_{0,y} = \Delta\Theta^2_{0,y} + 1 * \delta_y = 0.07733695992$

$D^1_{x2,h} = (-0.01333293759 + 0)/2 = 0.00666646879$

$D^1_{x1,h} = (-0.01141871967 + -0.01141871967)/2 = -0.01141871967$

$D^1_{0,h} = (0.00191421792 + -0.01141871967)/2 = -0.00475225087$

$D^2_{h,y} = (0.04252248982 + -0.25365575374)/2 = -0.10556663196$

$D^2_{0,y} = (0.07733695992 + -0.46133152004)/2 = -0.19199728006$

Update theta:

$\Theta^1_{0,h} = 0.1 - 0.3 * D^1_{0,h} + 0 = 0.10142567526$

$\Theta^1_{x1,h} = 0.1 - 0.3 * D^1_{x1,h} + 0 = 0.1034256159$

$\Theta^1_{x2,h} = 0.1 - 0.3 * D^1_{x2,h} + 0 = 0.10199994063$

$\Theta^2_{0,y} = 0.1 - 0.3 * D^2_{0,y} + 0 = 0.15759918401$

$\Theta^2_{h,y} = 0.1 - 0.3 * D^2_{h,y} + 0 = 0.13166998958$

3. (1). The range of the TANH function is [-1,1] and that of the sigmoid function is [0,1]. TANH has stronger gradients: since data is centered around 0, the derivatives are higher.

3.(2). TANH(x) = $\frac{SINH}{COSH} = \frac{1-e^{-2x}}{1+e^{2x}}$

TANH(x) +1 = $\frac{1-e^{-2x}}{1+e^{2x}} + \frac{1+e^{-2x}}{1+e^{2x}} = \frac{2}{1+e^{2x}} = 2\left(\frac{1}{1+e^{2x}}\right) = 2\sigma(2x)$

$\sigma(x)$ = $\frac{\text{TANH}(0.5x) + 1}{2}$

PartII.

1.

Which classier is better? Write 2-3 sentences justifying your answer, discussing the results you Obtained.

SVM seems to be better because it achieves higher scores in all of the metrics of precision, accuracy, and recall in the testing data.

TEXT20NEWS_ACCURACY_TEST_NB: 0.819437068508

TEXT20NEWS_ACCURACY_TEST_SVM: 0.846255974509

TEXT20NEWS_PRECISION_TEST_NB: 0.841965329356

TEXT20NEWS_PRECISION_TEST_SVM: 0.848552735505

TEXT20NEWS_RECALL_TEST_NB: 0.803871775123

TEXT20NEWS_RECALL_TEST_SVM: 0.839373845703

2. Report your optimal learning rate, regularization parameter, and the maximum training performance you obtained in your PDF writeup and README.

NEURALNET_ACCURACY_TRAIN: 0.9652

NEURALNET_OPTIMAL_LEARNINGRATE: 2.0

NEURALNET_LAMBDA: 0.0001

NEURALNET_OPTEPOCH: 750