Charles Zheng CME 323 HW 3

1. a. Consider a "broom" graph with nodes $1, \cdots, n$ with edges $(1, 2), \ldots, (n_1 - 1, n_1)$ and edges (n_1, i) for all $n_1 < i < n$, and finally an edge $(n-1, n)$. Then the worst-case k is $k = n_1 + 1$ and the worst-case IDs for this case are where $ID(i) = n - i + 1$ for $i = 1, \ldots, n$. Then what will happen is that in the first k iterations, each node except for node n will carry the ID of node 1, and all nodes $i \geq k$ (other than node n) will be of distance exactly k from node 1. Hence $n - k$ nodes will be added to the graph. The algorithm terminates on the $(k+1)$st iteration when the message from node $(n-1)$ carrying the ID of node 1 reaches node n.

b. For large (and even) n_1, the best ID distribution would be to have $ID(n_1/2) = n$, and the other ids so that $ID(i) > ID(j)$ for all i satisfying $i = n_1/2 + 2\ell k^*$ for integer ℓ and $i \leq n_1$ and all j not satisfying those conditions, where $k^* = \sqrt{\frac{3}{2}n_1}$. Here the optimal value of k is also k^*. In that case the total number of unique IDs at step $k = k^*$ is $\sqrt{\frac{2}{3}n_1}$, and the corresponding partitions are arranged along the line $1, 2, \ldots, n_1, n-1, n$. In the $k+1$th iterate, the maximal ID travels a distance of 1 to reach the edge of another partition, in the $k+2$nd it travels a distance of k^* to reach the node with the original ID of that partition, and in the $k+3$rd the maximal ID has completely replaced the ID of that partition. This happens in both left and right directions. In general, it takes 3 iterations for the maximal ID to spread a distance of $2k^* + 1$ in either direction. The maximum distance of the central node $n_1/2$ any other node is $n_1/2$. Hence it takes about $k^* + n_1/4k^* = n_1(\sqrt{3/2} + \sqrt{1/6})$ iterations.

c.

d.

2. Let w^* be an optimal weight vector with $||w^*|| = 1$ and $\min |x^T w^*| \geq \delta$. Let $w^{(0)} = 0$ and $w^{(k)}$ denote w after the kth mistake, and $x^{(k)}$ be the feature of the kth mistake, so that

$$\text{sign}(w^{(k)}x^{(k)}) = -\text{sign}(w^* x^{(k)})$$

and

$$w^{(k+1)} = w^{(k)} - \text{sign}(w^{(k)}x^{(k)})w^{(k)}$$

Then we have

$$||w^{(k+1)}||^2 = ||w^{(k)} - \text{sign}(w^{(k)}x^{(k)})x^{(k)}||^2 = ||w^{(k)}||^2 + ||x^{(k)}||^2 - |w^{(k)}x^{(k)}| \leq ||w^{(k)}|| + R^2$$

and also

$$\langle w^{(k+1)}, w^* \rangle = \langle w^{(k)} + \text{sign}(w^* x^{(k)}) x^{(k)}, w^* \rangle = \langle w^{(k)}, w^* \rangle + |\langle x^{(k)}, w^* \rangle| \geq \delta$$

These two facts imply by induction that

$$||w^{(k)}|| \leq R\sqrt{k}$$

and

$$\langle w^{(k)}, w^* \rangle \geq \delta k$$

But now observe that

$$1 \geq \cos(w^{(k)}, w^*) = \frac{\langle w^{(k)}, w^* \rangle}{||w^*||||w||} \geq \frac{\delta k}{R\sqrt{k}} = \frac{\delta \sqrt{k}}{R}$$

Hence we get

$$k \leq \frac{R^2}{\delta^2}$$

This means that the number of mistakes is bounded by $\frac{R^2}{\delta^2}$, since otherwise we would arrive at a contradiction.

3.

4. a. The communication cost of the shuffle is $O(n \log n)$ and the communication cost of the reduce is $O(k)$.

b. The total communication cost is $O(Tk)$ and the total time it takes to reduce + broadcast is $O(T \log k)$.