| | |
|---|---|
| **Machine Learning** | **Spring 2017** |

# Solution : Homework 3

*Lecturer: Yang Yang* — *Homework taker: Li Xu*

**Due Time:** May 19

**Problem 1.**

**Solution:**

*Answer for problem (1):*

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial z_m}\sigma'(a_1)\prod_{k=2}^{m}\sigma'(a_k)w_k$$

*Answer for problem (2):*

(a) For sigmoid function $\sigma$, we have:

$$\sigma'(x) = \sigma(x)(1-\sigma(x)) \leq \frac{1}{4}$$

Thus:

$$|w_j\sigma(a_j)| < \frac{1}{4} < 1$$

So when $m$ is large, $\prod_{m}^{k=2}\sigma'(a_k)w_k$ tends to be 0.

(b) Even if we have a large $w$, $\sigma'(a)$ is still very small as

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

*Answer for problem (3):*

As long as $a > 0$, $\sigma'(a) = 1$. Thus there is no vanishing problem. ■

**Problem 2.**

**Solution:**

We first prove $VC(h) \geq n+1$, that is to say, we first find a set of size $n+1$ that h can shatter. It's easy to construct. For example, we let the set equals to

$$(0,0,...,0)^T, (1,0,...,0)^T, ..., (n,0,...,0)^T$$

and $(0,0,...,0)$ is in class 1 the others are in class 2. If we let $a = (1,0,...,0)^T$ and $b = 0$, $h(x)$ then can shatter this set.

Next we prove $VC(h) \leq n+1$. There is no set with size $n+2$ that h can shatter. For a set of n + 2 points in $\mathbb{R}^n$ can be partitioned into two disjoint subsets $S_1$ and $S_2$ such that their convex hulls intersect. Thus it cannot be shattered linearly.

■

**Problem 3.**

**Solution:** In this data set, Relief is used for feature chosen. I chose top five features for each pair. For D and O the chosen features are [6,8,9,10,13], and for O and X are [8,9,10,11,13].

The number of nodes in hidden layer is set to 6 and the number of iterations is set to 200. learning rate is set to 0.005. Sigmoid function is used both in hidden layer and output layer. Other values are initialed randomly between 0 to 0.2.

The final accuracy rate between D and O is 98.93% and 98.92%between O and X. ■