# Homework 3

**Solution 1.**

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

2. (a) The derivative of sigmoid function reaches a maximum at $\sigma'(0) = \frac{1}{4}$. Since $|w_j| < 1$, we have $|w_j \sigma'(a_j)| < \frac{1}{4} < 1$.

   (b) To avoid the vanishing gradient problem we need $|w\sigma'(a)| \geq 1$. But, the $\sigma'(a)$ term also depends on $w : \sigma'(a) = \sigma(wz + b)$. If we make $w$ large we tend to make $wz + b$ very large, and $\sigma'(a)$ very small.

3. As long as $a > 0$, $\sigma'(a) = 1$. So we dont have the issue as in $2(b)$.

**Solution 2.**

We first show $H$ can shatter $n + 1$ points. Let $S = x_{i\,i=0}^{\,n}$ and $y_i \in 1, 1$ be the label of $x_i$ . If we can place $S$ such that $y_i(a^T x_i + b)geq0$ holds for all $y_i$ , then $S$ can be shattered by $H$. Let $x_0 = 0$ and $x_i$ be the unit vector on the $i$-th coordinate. Take $b = y_0/2$ and $a_i = y_i$ . Then

$$\begin{aligned}
y_0\dot{(}0 + b) &= \frac{1}{2}y_0^2 \geq 0 \\
y_1\dot{(}y_1 + b) &= y_1^2 + \frac{1}{2}y_0 y_1 \geq 0 \\
&\vdots \\
y_n\dot{(}y_n + b) &= y_n^2 + \frac{1}{2}y_0 y_n \geq 0
\end{aligned}$$

always hold. Therefore $\text{VCdim}(H) \geq n + 1$.

Now let $S$ contain $n+2$ points, we show $H$ cannot shatter $S$. Let $P = \{x : a^T x + b \geq 0\}$ be the halfspace defined by $h \in H$. Notice that $S \subseteq P \Rightarrow \textbf{conv}(S) \subseteq P$, since

$$a^T(\sum_{i=1}^{k} \alpha_i x_i) + b = \sum_{i=1}^{k} \alpha_i(a^T x_i + b) \geq 0$$

Similar for the opposite halfspace $P^c$ . Suppose $H$ can shatter $S$. Now $H$ can separate any disjoint subsets $S_1$ and $S_2$ such that $S_1 \subseteq P$ and $S_2 \subseteq P^c$ . By the claim above, this implies $\mathbf{conv}(S_1) \subseteq P$ and $\mathbf{conv}(S_2) \subseteq P^c$. However by Radons theorem there exist $S_1$ and $S_2$ whose convex hulls intersect. This is a contradiction. Hence $\mathrm{VCdim}(H) \leq n + 1$.