Adam Cameron

Assignment 1

1)

(final draft)

- (i) $n=1, k=1, v=\begin{pmatrix}1\\-1\end{pmatrix}, \sigma \doteq \text{ReLU}, w, u \in \mathbb{R}^d$

$$W=\begin{pmatrix}\vec{w}^T\\ \vec{u}^T\end{pmatrix} \in \mathbb{R}^{2k\times d} = \begin{pmatrix}\vec{w}^T\\ \vec{u}^T\end{pmatrix} \in \mathbb{R}^{2\times d}$$

$$g_w(x) = v^T\sigma(Wx), \quad Wx = \begin{pmatrix}\langle \vec{w}, \vec{x}\rangle\\ \langle \vec{u}, \vec{x}\rangle\end{pmatrix}$$

$$\langle \vec{w}, \vec{x}\rangle \doteq w^T x, \quad \langle \vec{u}, \vec{x}\rangle = u^T x,$$

$$g_w(x) = (1\ {-1})\,\sigma\begin{pmatrix}w^T x\\ u^T x\end{pmatrix}$$

$$= (1\ {-1})\begin{pmatrix}\sigma\langle w, x\rangle\\ \sigma\langle u, x\rangle\end{pmatrix}$$

$$= \sigma\langle w, x\rangle - \sigma\langle u, x\rangle \therefore$$

- (ii) (final draft)

$n=1, k=1, v=\begin{pmatrix}1\\-1\end{pmatrix}, y_1=-1$

$$L_S(W, \vec{v}) = L_S(w, v)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \max(1 - y_i g_w(x_i), 0)$$

Assignment 1

$$= \frac{1}{1} \sum_{i=1}^{n} \max(1 - y_i g_w(x_1), 0)$$

$$= \max(1 - y_1 g_w(x_1), 0)$$

$$= \max(1 - (-1)(\sigma(\langle w, x \rangle) - \sigma(\langle u, x \rangle)), 0)$$

$$= \max(1 + ((\sigma\langle w, x \rangle) - \sigma(\langle u, x \rangle)), 0) \therefore$$

$= \frac{1}{1} \sum_{i=1}^{1} \max(1 - y_1 g_w(x_1), 0)$

$= \max(1 - y_1 g_w(x_1), 0)$

$= \max(1 - (-1)((\sigma(\langle w, x\rangle) - \sigma(\langle u, x\rangle)), 0)$

$= \max(1 + (\sigma(\langle w, x\rangle) - \sigma(\langle u, x\rangle)), 0)$ $\therefore$

2)

set w to $\frac{w_1 + w_2}{2}$, u to $\frac{u_1 + u_2}{2}$,

$w_1 = w_2 = u_1 = x$, $u_2 = -x$,

(plug in)

$L_S\left(\left(\frac{w_1 + w_2}{2}\right), \left(\frac{u_1 + u_2}{2}\right)\right)$

(hinge loss) $= \max\left(1 + \left(\mathrm{ReLU}\left(\left\langle \frac{w_1 + w_2}{2}, x\right\rangle\right) - \mathrm{ReLU}\left(\left\langle \frac{u_1 + u_2}{2}, x\right\rangle\right)\right), 0\right)$

(hinge loss) $= \max\left(1 + \left(\mathrm{ReLU}\left(\left\langle \frac{x + x}{2}, x\right\rangle\right) - \cancel{\mathrm{ReLU}\left(\frac{x - x}{2}\right)}\right.\right.$

Assignment 1

2)

$$\text{ReLU}\left(\left\langle \frac{x-x}{2}, x \right\rangle\right)\Big), 0\Big)$$

$$= \max(1 + (\text{ReLU}(x \cdot x) - \text{ReLU}(0 \cdot x)), 0)$$

$$= \max(1 + (\text{ReLU}(\|x\|^2) - \text{ReLU}(0)), 0)$$

$$= \max(1 + \|x\|^2, 0)$$

$$= 1 + \|x\|^2;$$

- $$L_S(w_1, u_1) = \max(1 + (\sigma(w_1, x) - \sigma(u_1, x)), 0)$$

$$= \max(1 + (\sigma(\langle x, x \rangle) - \sigma(\langle x, x \rangle)), 0)$$

$$= \max(1 + \|x\|^2, 0) = 1 + \|x\|^2;$$

- $$L_S(w_2, u_2) = \max(1 + (\sigma(\langle (w_2), x \rangle) - \sigma(\langle (u_2), x \rangle)), 0)$$

$$= \max(1 + (\sigma(\langle (x), x \rangle) - \sigma(\langle (-x, x) \rangle)), 0)$$

$$= \max(1 + (\sigma(\|x\|^2) - \sigma(-\|x\|^2)), 0)$$

$$= \max(1 + (\|x\|^2 - 0), 0) = \max(1 + \|x\|^2, 0)$$

$= 1 + \|x\|^2;$

- $\frac{1}{2}\left(L_S(w_1, v_1) + L_S(w_2, v_2)\right) = \frac{1}{2}\left((1) + (1 + \|x\|^2)\right)$

$= \frac{1}{2}(2 + \|x\|^2) = 1 + \frac{\|x\|^2}{2};$

- $0 < \|x\|^2 < 1,$

$1 + \|x\|^2 > 1 + \frac{\|x\|^2}{2},$

$L_S\left(\frac{w_1 + w_2}{2}, \frac{v_1 + v_2}{2}\right) > \frac{1}{2}\left(L_S(w_1, v_1) + L_S(w_2, v_2)\right) \therefore$

(Chapter 2 (2.2)) 3) $[\alpha > 0, [z_\alpha = (1 - \alpha)x + \alpha y]$

$f(z_\alpha) = f(x) + (\nabla f(x))^T (z_\alpha - x) + O(|z_\alpha - x|^2)$

gradient Transpose

Effectively, this problem is asking to prove Theorem 2.3 (1) (2.12)

Start $f((1 - \alpha)x + \alpha y) \leq (1 - \alpha)f(x) + \alpha f(y),$

Assignment 1

plug in

$$f(z_\alpha) \le (1-\alpha) f(x) + \alpha f(y),$$

$$f(z_\alpha) - f(x) \le \alpha (f(y) - f(x)),$$

$$z_\alpha = (1-\alpha)x + \alpha y, \quad z_\alpha - x = \alpha (y-x),$$

$$\frac{f(z_\alpha) - f(x)}{z_\alpha - x} \le \frac{\alpha (f(y) - f(x))}{z_\alpha - x},$$

$$\frac{f(z_\alpha) - f(x)}{z_\alpha - x} = f'(x), \quad \frac{\alpha (f(y) - f(x))}{z_\alpha - x} = \frac{f(y)-f(x)}{y-x},$$

$$f'(x) \le \frac{f(y) - f(x)}{y-x},$$

left side

$$f(y) - f(x) \ge f'(x)(y-x),$$

$$f(y) - f(x) \ge (\nabla f(x))^T (y-x),$$

$$f(y) \ge f(x) + (\nabla f(x))^T (y-x) \therefore$$