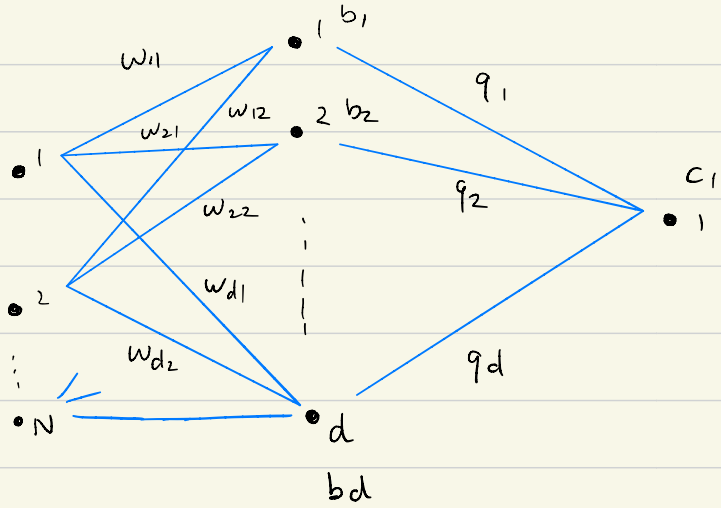
Singular Learning Theory 8

This note is a continuation of "Singular Learning Theory 5" (sdl5), where a particular class of models based on feedforward ReLU networks were defined. In this note we theoretically derive the RLCT of these models paired with the symmetric true distributions.

Recall from (6.1) of (sdl5) that we consider a two-layer ReLU network



(1.1)

and the corresponding function

$$f : \mathbb{R}^2 \times W \longrightarrow \mathbb{R} \tag{1.2}$$

$$f(x,w) = c_1 + \sum_{i=1}^{d} q_i \, \mathrm{ReLU}(\langle w_{i\bullet}, x \rangle + b_i)$$

where $w = (w_{1\bullet}, \ldots, w_{d\bullet}, b_1, \ldots, b_d, q_1, \ldots, q_d, c_1)$, and $W \subseteq \mathbb{R}^{41}$ is compact. We denote by H_i the following subspace

$$H_i = \{ x \in \mathbb{R}^2 \mid \langle w_{i\bullet}, x \rangle + b_i = 0 \} \tag{1.3}$$

We assume the distribution $q(y|x)$ is realisable, and that a prior φ is fixed which is nonzero on the set of true parameters.

The model class is $p(y|x,w)$ where (see "Fisher for feedforward")

$$p(y|x,w) := \frac{1}{(2\pi)^{1/2}} \exp\left(-\tfrac{1}{2}\| y - f(x,w)\|^2\right) \tag{2.1}$$

and the Kullback–Leibler distance to the true distribution is

$$K(w) = \int q(y|x)\, q(x) \log \frac{q(y|x)}{p(y|x,w)}\, dx\, dy \tag{2.2}$$

where $q(x)$ is a chosen probability density function on $\mathbb{R}^2$. We are first of all interested in the set of true parameters

$$W_0 = \{w \in W \mid K(w) = 0\} \tag{2.3}$$

which is nonempty by hypothesis. The true distribution we consider here is inspired by, but different from, that in (sdl5). The reason for the change is that the $\mathbb{Z}/m\mathbb{Z}$ symmetry no longer appears central.

Hypothesis I we suppose given an integer $1 \le m \le d$ and lines $H_1^{(0)}, \dots, H_m^{(0)} \subseteq \mathbb{R}^2$ together with chosen $w_{i\bullet}^{(0)} \in \mathbb{R}^2$ and $b_i^{(0)} \in \mathbb{R}$ for $1 \le i \le d$ such that $H_i^{(0)}$ is the line determined by $w_{i\bullet}^{(0)}, b_i^{(0)}$ as in (1.3) (note that since H_i is a line, this means $w_{i\bullet}^{(0)} \neq 0$ for $1 \le i \le d$) and $H_i^{(0)} \neq H_j^{(0)}$ whenever $i \neq j$. Then

$$q(y|x) := p(y|x, w^{(0)}) \tag{2.4}$$

where $w^{(0)}$ is the weight vector (assumed to be in the interior of W)

$$w^{(0)} = (w_{1\bullet}^{(0)}, \dots, w_{m\bullet}^{(0)}, \underbrace{0, \dots, 0}_{d-m}, b_1^{(0)}, \dots, b_m^{(0)}, \underbrace{0, \dots, 0}_{d-m}, \underbrace{1, \dots, 1}_{m}, \underbrace{0, \dots, 0}_{d-m}, 0)$$

Explicitly

$$q(y|x) = \frac{1}{(2\pi)^{1/2}} \exp\left(-\tfrac{1}{2} \| y - \sum_{i=1}^{m} \mathrm{ReLU}(\langle w_{i\bullet}^{(0)}, x \rangle + b_i^{(0)}) \|^2 \right) \qquad (3.1)$$

so the true distribution is a generic model determined by a ReLU network of depth two with m hidden nodes (the values of c and the q_i not being important).

Definition Given $w, w' \in W$ we define

$$K(w', w) = \int p(y|x,w')\, q(x) \log \frac{p(y|x,w')}{p(y|x,w)}\, dx dy$$

so that $K(w) = K(w^{(0)}, w)$.

We calculate

$$K(w',w) = \int q(x) \frac{1}{(2\pi)^{1/2}} \exp\left(-\tfrac{1}{2} \| y - f(x,w') \|^2 \right) \cdot \log \frac{\exp(-\frac{1}{2}\| y - f(x,w')\|^2)}{\exp(-\frac{1}{2}\| y - f(x,w)\|^2)}\, dx dy$$

$$= \frac{1}{(2\pi)^{1/2}} \int q(x) \exp\left(-\tfrac{1}{2} \| y - f(x,w') \|^2 \right) \cdot \tfrac{1}{2} \left\{ \| y - f(x,w)\|^2 - \| y - f(x,w')\|^2 \right\} dx dy$$

Fix $x \in \mathbb{R}^2$ and let $u = y - f(x,w')$, $a = f(x,w) - f(x,w')$ so that

$$K(w',w) = \frac{1}{2(2\pi)^{1/2}} \int q(x)\, K(w',w,x)\, dx \qquad u^2 - 2au + a^2 - u^2$$

$$K(w',w,x) = \int \exp\left(-\tfrac{1}{2} u^2 \right) \cdot \left\{ (u-a)^2 - u^2 \right\} du$$

$$= a^2 \underbrace{\int \exp(-\tfrac{1}{2}u^2)\, du}_{\sqrt{2\pi}} - 2a \underbrace{\int u \exp(-\tfrac{1}{2}u^2)\, du}_{0}$$

Hence by standard techniques (see "Fisher for feedforward")

$$K(w', w, x) = \sqrt{2\pi}\, a^2 \tag{4.1}$$

Hence

$$K(w', w) = \frac{1}{2}\int q(x) \,\|\, f(x,w) - f(x,w') \,\|^2 \, dx \tag{4.2}$$

Note that since $f(x,w)$ is piecewise-linear, on partitions determined by at most d hyperplanes, to determine if $f(x,w) = f(x,w')$ on all of $\mathbb{R}^2$ it suffices to compare their values in some compact neighbourhood of 0. Without attempting to make this precise (at the moment) we assume $q(x)$ is positive in a big enough neighbourhood of 0 so that this comparison suffices for all $w \in W$.

Hypothesis II for all $w \in W$, $q(x)$ is such that $K(w) = 0$ if and only if

$$f(x,w) = f(x, w^{(0)}) \quad \text{for all } x \in \mathbb{R}^2.$$

So (4.2) should be read as an assertion that there are many natural choices for $q(x)$ satisfying Hypothesis II. Hence

$$W_0 = \{ w \in W \mid f(x,w) = f(x, w^{(0)}) \text{ for all } x \in \mathbb{R}^2 \} \tag{4.3}$$

which reduces the problem of classifying true parameters to the problem of classifying weights $w \in W$ up to functional equivalence of the corresponding ReLU network. For this we use [PL] as a starting point.

Remark What we have said so far generalises to ReLU networks of arbitrary depth computing functions $\mathbb{R}^N \longrightarrow \mathbb{R}^M$.

Fold sets following the notation of [PL] we define the **fold set** of a continuous piecewise-linear function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ by

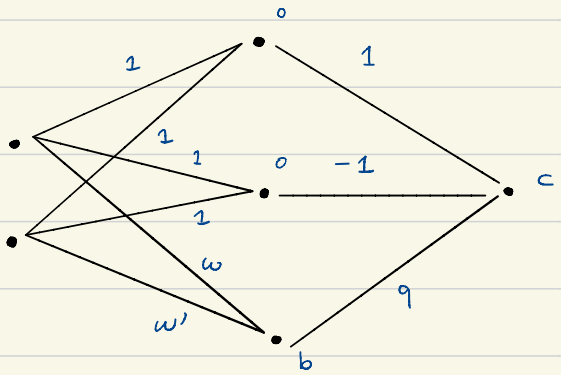
$$\mathcal{F}(f) = \{x \in \mathbb{R}^2 \mid f \text{ is not differentiable at } x\}$$

Def^n Given a piecewise-linear $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ let $\{U_\alpha\}_{\alpha \in \Lambda}$ be the domains in $\mathbb{R}^2$ such that $\bigcup_\alpha U_\alpha = \mathbb{R}^2 \setminus \mathcal{F}(f)$. Then for each α there are unique $w^\alpha \in \mathbb{R}^2$ and $b^\alpha \in \mathbb{R}$ such that $f(x) = \langle w^\alpha, x \rangle + b^\alpha$ for all $\alpha \in \bigcup \alpha$.

Example By hypothesis $\mathcal{F}(\, f(x, w^{(0)}) \,) = \bigcup_{i=1}^{m} H_i^{(0)}$.

Note that in general, for $w \in W$ some of the H_i may be empty, or all of $\mathbb{R}^2$, and even if they are all lines the fold set may be a proper superset of $\bigcup_i H_i$, as the following example shows.

Example Consider the network $f(x, w) = c + q\,\mathrm{ReLU}(w x_1 + w' x_2 + b)$



whose fold set is, assuming $q \neq 0$, the line $w x_1 + w' x_2 + b = 0$.

Next we consider the simplest case, where m (the number of hyperplanes determining the true distribution $q(y|x)$) is equal to cl (the number of hyperplanes "available" in the model).

Case m=d Suppose $f(x,w) = f(x,w^{(0)})$ as functions on $\mathbb{R}^2$. Then

$$\mathcal{F}(f(x,w)) = H_1^{(0)} \cup \cdots \cup H_m^{(0)}$$

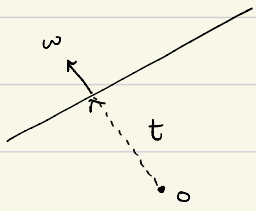
which implies that (writing w, q, b, c for the components of w) none of the q_i are zero, and none of $w_{i\bullet}$ are zero for $1 \le i \le m$. Further there is a permutation $\beta \in S_3$ such that $H_i = H^{(0)}_{\beta i}$ for $1 \le i \le m$.

Lemma For $w, w' \in \mathbb{R}^2$ and $b, b' \in \mathbb{R}$ we set

$$H = \{x \in \mathbb{R}^2 \mid \langle w, x\rangle + b = 0\}$$
$$H' = \{x \in \mathbb{R}^2 \mid \langle w', x\rangle + b' = 0\}.$$

Suppose H, H' are both lines. Then $H = H'$ if and only if there exists $\lambda \in \mathbb{R}\setminus\{0\}$ such that $w = \lambda w'$, $b = \lambda b'$.

Proof One direction is clear. Since $w \neq 0$ we can find unique t s.t. $\langle w, t\rangle = -b$, and similarly $\langle w', t'\rangle = -b'$, with t a multiple of w and t' a multiple of w'. Then $t \in H$, $t' \in H'$ and



$$\langle w, x - t\rangle = 0 \iff x \in H$$
$$\iff x \in H'$$
$$\iff \langle w', x - t'\rangle = 0$$

Hence $\langle w', t - t'\rangle = 0$ and $\langle w, t - t'\rangle = 0$, so if w, w' are linearly independent $t = t'$, which is a contradiction since $t \in \text{span}(w)$, $t' \in \text{span}(w')$. Hence $\{w, w'\}$ is linearly dependent, say $w = \lambda w'$. Then $-b = \langle w, t\rangle = \langle w, t'\rangle = \lambda\langle w', t'\rangle = -\lambda b'$ so $b = \lambda b'$. □

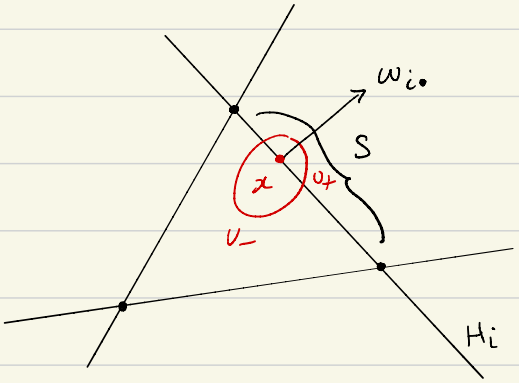
It follows that for $1 \le i \le m$ there exists a (unique) $\lambda_i \in \mathbb{R}_{>0}$ and $\epsilon_i \in \mathbb{Z}_2$ s.t.

$$\begin{aligned} w_{i\bullet} &= \lambda_i (-1)^{\epsilon_i} w^{(0)}_{6i,\bullet} \\ b_i &= \lambda_i (-1)^{\epsilon_i} b^{(0)}_{6i} \end{aligned} \tag{7.1}$$

Hence

$$\begin{aligned} f(x,w) &= c_1 + \sum_{i=1}^m q_i \, \mathrm{ReLU}\left(\langle w_{i\bullet}, x \rangle + b_i \right) \\ &= c_1 + \sum_{i=1}^m q_i \, \mathrm{ReLU}\left(\lambda_i (-1)^{\epsilon_i} \{ \langle w^{(0)}_{6i}, x \rangle + b^{(0)}_{6i} \} \right) \\ &= c_1 + \sum_{i=1}^m q_i \lambda_i \, \mathrm{ReLU}\left((-1)^{\epsilon_i} \{ \langle w^{(0)}_{6i}, x \rangle + b^{(0)}_{6i} \} \right) \end{aligned} \tag{7.2}$$

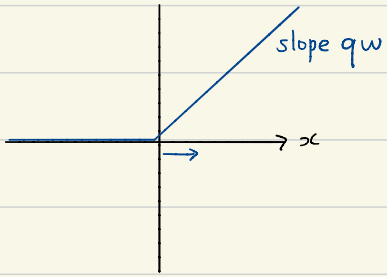
Using w we may assign an orientation to the normal bundle of the 1-strata of $\mathcal{F}(f(x,w))$ in $\mathbb{R}^2$, according to the direction of the velevant $w_{i\bullet}$. More precisely, let $x \in S \subseteq \mathcal{F}(f(x,w))$ be a point in a 1-strata S, which is contained in H_i. Then $w_{i\bullet}$ gives the direction in which the coresponding ReLU is "active".



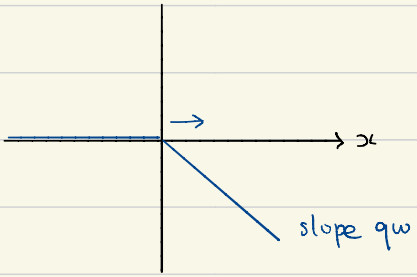
(7.3)

Picking $x \in U = U_+ \cup U_-$ small enough we have $f(x,w) = \langle w^\pm, x \rangle + b^\pm$ for $x \in U^\pm$ where $w^+ - w^- = q_i w_{i\bullet}$. The point here is that $w^+ - w^-$ may be inferred from the function $f(x,w)$ by looking at its values near x, and hence so may $q_i w_{i\bullet}$

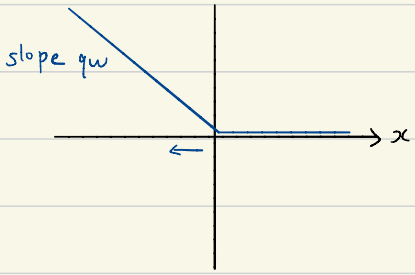
Example $f(x, q, w) = q\,\mathrm{ReLU}(wx)$ has a decision boundary at $x = 0$ with normal vector $+$ if $w > 0$ and $-$ if $w < 0$.



$w > 0, q > 0$



$w > 0, q < 0$



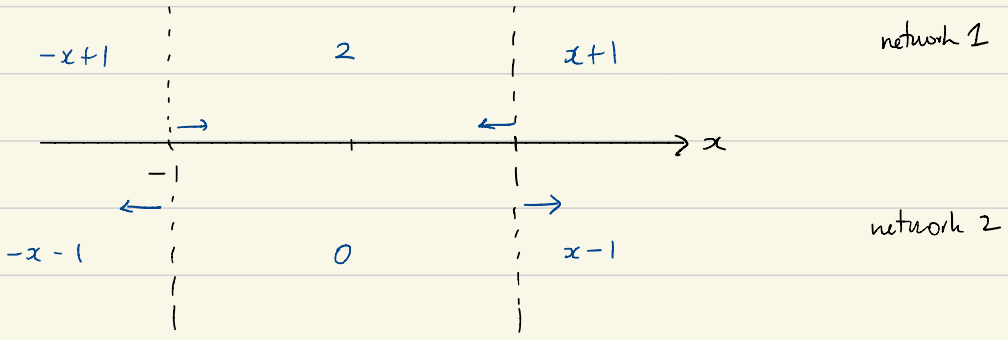
$w < 0, q > 0$

$w < 0, q < 0$

Example Consider the networks

$$\mathrm{ReLU}(x+1) + \mathrm{ReLU}(-x+1) \quad \text{network 1 (truth)}$$

$$= 2 + \mathrm{ReLU}(-x-1) + \mathrm{ReLU}(x-1) \quad \text{network 2.}$$



The line $\mathbb{R}$ is divided into regions $U_\alpha = (-\infty, -1)$, $U_\beta = (-1, 1)$, $U_\gamma = (1, \infty)$ and both networks have $w^\alpha = -1$, $w^\beta = 0$, $w^\gamma = 1$ but we see that $\epsilon_1 = \epsilon_2 = 1$, so this is possible.

This means that for each $1 \le i \le m$ we are in one of two cases:

(a) $\epsilon_i = 0$, so $w_{i\bullet}, w_{bi,\bullet}^{(0)}$ are the same direction, and $q_i w_{i\bullet} = q_i^{(0)} w_{bi,\bullet}^{(0)}$ so

$$w_{i\bullet} = \frac{q_i^{(0)}}{q_i} w_{bi,\bullet}^{(0)} \tag{9.1}$$

(b) $\epsilon_i = 1$ so $w_{i\bullet}, w_{bi,\bullet}^{(0)}$ are in opposite directions and $q_i w_{i\bullet} = -q_i^{(0)} w_{bi,\bullet}^{(0)}$ so

$$w_{i\bullet} = -\frac{q_i^{(0)}}{q_i} w_{bi,\bullet}^{(0)} \tag{9.2}$$

Together this shows that for $1 \le i \le m$ (noting $q_i^{(0)} = 1$)

$$w_{i\bullet} = (-1)^{\epsilon_i} \frac{q_i^{(0)}}{q_i} w_{bi,\bullet}^{(0)} \tag{9.3}$$

Comparison with (7.1) shows that q_i must be positive, that $\lambda_i = 1/q_i$ and

$$b_i = (-1)^{\epsilon_i} \frac{q_i^{(0)}}{q_i} b_{bi}^{(0)} \tag{9.4}$$

As the example on the previous page shows, it is not necessarily the case that all ϵ_i are zero. But these may be nonzero only under special conditions:

Lemma Set $F = \{i \mid \epsilon_i = 1\}$. Then $\sum_{i \in F} w_{bi,\bullet}^{(0)} = 0$.

Proof With the notation of p.(5) let U_α be a domain and let δ_i^α be $+1$ if this domain is in the active region for the ReLU with vector $w_{i,\bullet}^{(0)}$ in the network $w^{(0)}$ computing $f(x, w^{(0)})$ (so $\langle w_{i,\bullet}^{(0)}, x \rangle + b_i^{(0)} > 0$), and zero otherwise. Let $\bar{\delta}_i^\alpha$ be $+1$ if $\delta_i^\alpha = 0$ and 0 otherwise. Then

$$\sum_i \delta_{bi}^\alpha w_{bi,\bullet}^{(0)} = \sum_{i \notin F} \delta_{bi}^\alpha q_i w_{i,\bullet} + \sum_{i \in F} \bar{\delta}_{bi}^\alpha q_i w_{i,\bullet} \tag{9.5}$$

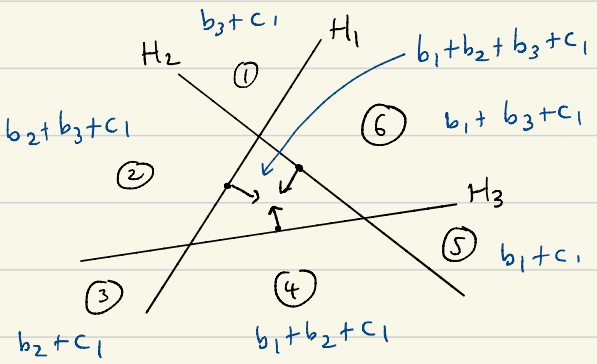
Hence

$$\sum_{i\notin F} \delta^{\alpha}_{\beta i} w^{(0)}_{\beta i,\bullet} + \sum_{i\in F} \delta^{\alpha}_{\beta i} w^{(0)}_{\beta i,\bullet}$$

$$= \sum_{i\notin F} \delta^{\alpha}_{\beta i} w^{(0)}_{\beta i,\bullet} + \sum_{i\in F} \bar{\delta}^{\alpha}_{\beta i} (-w^{(0)}_{\beta i,\bullet})$$

which implies $\sum_{i\in F} \left[\delta^{\alpha}_{\beta i} + \bar{\delta}^{\alpha}_{\beta i} \right] w^{(0)}_{\beta i,\bullet} = 0$ as claimed. □

Example Consider $f(x,w)$ determing hyperplanes as shown, with vectors $w_{1\bullet}, w_{2\bullet}, w_{3\bullet}$. In all six regions some ReLU is active



Suppose the function $f(x,w)$ is known, together with the q_i, $w_{i\bullet}$ and b_i. Then c_1 may be found by evaluating

$$c_1 = f(x,w) - \sum_{i=1}^{m} q_i \, \mathrm{ReLU}\left(\langle w_{i\bullet}, x \rangle + b_i \right)$$

at any point of $\mathbb{R}^2$. One way of putting this is that if you know the decision boundaries and orientations, you know how q_w and q_b vary across boundaries and hence for any $x \in \mathbb{R}^2$ you can determine $\sum_j q_j b_j + c_1$ (j ranging over active indices) by examining $f(x,w)$ (i.e. this is b^{α} in the notation of p. ⑰) and then subtract $\sum_j q_j b_j$.

Lemma Set $F = \{i \mid \epsilon_i = 1\}$. Then $\sum_{i \in F} b^{(0)}_{\beta i} = c_1$.

Proof By the same argument as given for the weights

$$\sum_i \delta^{\alpha}_{\beta i} b^{(0)}_{\beta i} = c_1 + \sum_{i \notin F} \delta^{\alpha}_{\beta i} q_i b_i + \sum_{i \in F} \bar{\delta}^{\alpha}_{\beta i} q_i b_i$$

$$= c_1 + \sum_{i \notin F} \delta^{\alpha}_{\beta i} b^{(0)}_{\beta i} + \sum_{i \in F} \bar{\delta}^{\alpha}_{\beta i} (-b^{(0)}_{\beta i})$$

Hence $\sum_{i \in F} \left[\delta^{\alpha}_{\beta i} + \bar{\delta}^{\alpha}_{\beta i} \right] b^{(0)}_{\beta i} = c_1$ as claimed. □

Note that the lemma applies even if F is empty, to show $c_1 = 0$.

Set $X = \{(\lambda, q) \in \mathbb{R}_{>0} \mid \lambda q = 1\}$ and let

$$\mathcal{P} = \left\{ \epsilon : \{1, \dots, m\} \to \mathbb{Z}_2 \;\middle|\; \sum_{\epsilon_i = 1} \omega^{(0)}_{i,\bullet} = 0 \right\}$$

Our convention is that the empty sum is zero, so $\epsilon \equiv 0$ is in $\mathcal{P}$.

Proposition In the case $m = d$ there is a bijection (S_m is the set of permutations)

$$\Psi : X^m \times S_m \times \mathcal{P} \xrightarrow{\simeq} W_0$$

$$\left((\lambda_i, q_i)_{i=1}^m, \beta, \epsilon \right) \longmapsto \Big(\left(\lambda_i (-1)^{\epsilon_{\beta i}} \omega^{(0)}_{\beta i, \bullet} \right)_{i=1}^m, \left(\lambda_i (-1)^{\epsilon_{\beta i}} b^{(0)}_{\beta i} \right)_{i=1}^m, (q_i)_{i=1}^m, \sum_{\epsilon_i = 1} b^{(0)}_i \Big) \tag{11.1}$$

Proof The function computed is (writing $\epsilon'_i = \epsilon_{bi}$)

$$\sum_{\epsilon_i=1} b_i^{(0)} + \sum_{i=1}^{m} \mathrm{ReLU}\left((-1)^{\epsilon'_i}\{\langle w_{bi,\bullet}^{(0)}, x\rangle + b_{bi}^{(0)}\}\right) \tag{12.1}$$

This has the same foldset as $f(x, w^{(0)})$ and so it suffices to show that for each open domain U_α both (11.2) and $f(x, w^{(0)})$ determine the same w^α, b^α. Let $\delta_i^\alpha, \bar{\delta}_i^\alpha$ be as on p. ⑨ (so they are referring to $f(x, w^{(0)})$). Then the "slope" w^α computed by (11.2) is ($F = \{i \mid \epsilon'_i = 1\}$)

$$\sum_{i \notin F} \delta_{bi}^\alpha w_{bi,\bullet}^{(0)} + \sum_{i \in F} \bar{\delta}_{bi}^\alpha (-w_{bi,\bullet}^{(0)}) \tag{12.2}$$

whereas the slope computed by $f(x, w^{(0)})$ is $\sum_i \delta_{bi}^\alpha w_{bi,\bullet}^{(0)}$. However by hypothesis $\sum_{i\in F}[\delta_{bi}^\alpha + \bar{\delta}_{bi}^\alpha] w_{bi,\bullet}^{(0)} = \sum_{i\in F} w_{bi,\bullet}^{(0)} = 0$ so these are equal. The "intercept" b^α computed by (11.2) is

$$\sum_{i\in F} b_{bi}^{(0)} + \sum_{i\notin F} \delta_{bi}^\alpha b_{bi}^{(0)} + \sum_{i\in F} \bar{\delta}_{bi}^\alpha (-b_{bi}^{(0)})$$

$$= \sum_{i\in F} \delta_{bi}^\alpha b_{bi}^{(0)} + \sum_{i\notin F} \delta_{bi}^\alpha b_{bi}^{(0)} \tag{12.3}$$

$$= \sum_i \delta_{bi}^\alpha b_{bi}^{(0)}$$

which is the intercept computed by $f(x, w^{(0)})$. Hence $f(x, w^{(0)}) = f(x, w)$, where w is the vector on the RHS of (11.1). Injectivity is clear, since the hyperplanes giving the decision boundaries are distinct so δ may be recovered, and thus also ϵ. Surjectivity follows from what we have said above. □

Definition For $\beta \in S_m$, $\epsilon \in \mathcal{P}$ let $W_0^{\beta,\epsilon} \subseteq W_0$ denote the image of Ψ restricted to $X^m \times \{\beta\} \times \{\epsilon\}$.

The subset $W_0^{\beta,\epsilon} \subseteq W_0$ is a submanifold and W_0 is the disjoint union over all β, ϵ of the $W_0^{\beta,\epsilon}$. Let $C^{\beta,\epsilon}$ be a compact neighbourhood of $W_0^{\beta,\epsilon}$ in W_0 which does not meet any other component $W_0^{\beta',\epsilon'}$. Recall the local RLCT of p.⑥ (sd14).

Definition Let $\lambda^{\beta,\epsilon}$ denote the learning coefficient of (p, q, φ) restricted to $C^{\beta,\epsilon}$.

Recall that W has dimension $3m+1$.

Proposition $\lambda^{\beta,\epsilon} = m + \frac{1}{2}$, for all β, ϵ.

Proof Since $W_0^{\beta,\epsilon}$ is a submanifold of dimension m [W, Remark 7.3] gives the inequality $\lambda^{\beta,\epsilon} \leq m + \frac{1}{2}$ directly. But for equality we need to say more. Set $\Omega = C^{\beta,\epsilon}$ and let $\mathcal{A}_\Omega$ be the sheaf of real analytic functions on Ω and $\mathcal{I}$ the ideal sheaf generated by K so that the structure sheaf of $W_0^{\beta,\epsilon}$ is $\mathcal{A}_\Omega / \mathcal{I}$. We claim that $W_0^{\beta,\epsilon}$ is a regular real-analytic variety. This follows from the $(\mathbb{R}_{>0})^m$-action on Ω being transitive on $W_0^{\beta,\epsilon}$ since the singular locus is closed (in whatever topology you like). Hence at all points of $W_0^{\beta,\epsilon}$ the RLCT is just half the codimension [L, Theorem 5.1]. □

Note that the structure sheaf on $W_0^{\beta,\epsilon}$ as a real-analytic variety is not determined by $W_0^{\beta,\epsilon}$ as a subset and it is the structure sheaf which determines which points of this subvariety are regular (at a regular point the RLCT is the codimension). Since we have not analysed K we cannot say much about $\mathcal{O}_{W_0^{\beta,\epsilon}}$ directly. However the singular locus is closed [H, Theorem 5.3] so a generic point of $W_0^{\beta,\epsilon}$ must be regular. Since $(\mathbb{R}_{>0})^m$ acts transitively no point is special, so they must all be generic hence regular.

Corollary The global RLCT of (p, q, φ) is $m + \frac{1}{2}$.

References

[W] S. Watanabe "Algebraic geometry and statistical learning"

[PL] M. Phuong, C. H. Lampert "Functional vs. parametric equivalence of ReLU networks" ICLR 2020.

[L] S. Lin "Useful facts about RLCT" 2012.

[H] R. Hartshorne "Algebraic geometry"