A. Appendix - Proofs

Proof of Lemma 3.1. By definition 3.3, it suffices to prove \( \forall i \in [n], T_i(x) \leq T_i(y) \). Let \( w_{ij} \) be the \( i \)th row and \( j \)th column component of \( W \):

\[
T_i(x) - T_i(y) = (\sum_{j \in [m]} w_{ij} x_j + b_i) - (\sum_{j \in [m]} w_{ij} y_j + b_i) = \sum_{j \in [m]} w_{ij} (x_j - y_j) \leq 0
\]

Proof of Theorem 3.1. DNN\(^+\) is a composition of monotone functions: layers of non-decreasing activation functions \( \Phi \) and affine transformation with all non-negative weights \( T : \mathbb{R}^m \to \mathbb{R}^n \) (Lemma 3.1). By closure of monotone function under compositionality, we have that DNN\(^+\) is an order-preserving monotone function. \( \square \)

Proof of Corollary 3.1.1. Without loss of generality, we assume the feature index pair \( i, j \in [n] \) satisfy \( a_i a_j < 0 \). Let \( d = (d_1, \ldots, d_i, \ldots, d_j, \ldots, d_n) \) be a point on the segment, i.e., \( a d + b = 0 \). Now we construct three points \( A, B, C \) with \( \epsilon > 0 \).\(^4\)

\[
\begin{align*}
A &= (d_1, \ldots, d_i - \epsilon, \ldots, d_j, \ldots, d_n) \\
B &= (d_1, \ldots, d_i + \epsilon, \ldots, d_j, \ldots, d_n) \\
C &= (d_1, \ldots, d_i + \epsilon, \ldots, d_j - 2 \frac{a_i}{a_j} \epsilon, \ldots, d_n)
\end{align*}
\]

First since \( a_i a_j < 0 \) and \( \epsilon > 0 \), we have \( 2 \frac{a_i}{a_j} \epsilon < 0 \), thus

\[
\begin{align*}
x_i^A &= x_i^B = x_i^C \implies A \preceq B \preceq C \\
F^+(A) &\leq F^+(B) \leq F^+(C)
\end{align*}
\]

Simultaneously, from \( a d + b = 0 \), we also have

\[
\begin{align*}
aA + b &= a d + b - a_i \epsilon = -a_i \epsilon \\
aB + b &= a d + b + a_i \epsilon = a_i \epsilon \\
aC + b &= a d + b + a_i \epsilon - 2 \frac{a_i}{a_j} a_j \epsilon = -a_i \epsilon
\end{align*}
\]

Then \( A, C \) must lie on the same side of \( L \) but different than \( B \), thus \( f(A) = f(C) \neq f(B) \). By the same logic as in Theorem 3.3, this contradicts with Eq 17; therefore, DNN\(^+\) cannot solve classification problems where the decision boundaries \( \{L\} \) have any segment \( L \) with a normal \( a = (a_1, \ldots, a_n) \) where \( \exists i \neq j \in [n], a_i a_j < 0 \). \( \square \)

Proof of Corollary 3.1.2. Without loss of generality, let’s assume region \( R_0 \in \{ R \} \) is a closed set. We denote all points in \( R_0 \) as a general form \( x = (x_1, \ldots, x_n) \). Consider any point \( B = (x_1^B, \ldots, x_n^B) \in R_0 \), we can always find two points \( A', C' \in \partial R \) that follow \( A' \preceq B \preceq C' \) with the following construction method:

\[
\begin{align*}
A' &= (\min(x_1), x_2^B, \ldots, x_n^B) \\
C' &= (\max(x_1), x_2^B, \ldots, x_n^B)
\end{align*}
\]

As we move \( \epsilon > 0 \) away from the boundary, we can further construct two points \( A, C \notin R_0 \) where

\[
\begin{align*}
A &= (\min(x_1) - \epsilon, x_2^B, \ldots, x_n^B) \\
C &= (\max(x_1) + \epsilon, x_2^B, \ldots, x_n^B)
\end{align*}
\]

Thus \( A \preceq B \preceq C \implies F^+(A) \leq F^+(B) \leq F^+(C) \), yet we have \( B \in R_0 \) while \( A, C \notin R_0 \implies f(A) = f(C) \neq f(B) \). By the same reasoning as in theorem 3.3, we have a contradiction thus a DNN\(^+\) cannot solve problems where the decision boundary forms a closed set. \( \square \)

Proof of Corollary 3.1.3. Without loss of generality, we assume \( R_0, R_1 \) are disconnected and belong to the same class, i.e., \( f(x) = c_1, \forall x \in R_0 \cup R_1 \), \( c_1 \) is a constant. Now consider any pair of points \( (A, B), A \in R_0, B \in R_1 \), the straight line segment \( AB \) that connects \( A \) and \( B \) must pass through another class by Definition 4.4. This means we must have point \( C \) on line \( AB \), but \( f(c) = c_2 \neq c_1 \) where \( c_2 \) is a constant. Next, we discuss the order relationship between \( A \) and \( B \):

**case 1** Exists such a pair \( A \preceq B \). Then since \( A, C, B \) are colinear and \( C \) is in between \( A \) and \( B \), we have \( A \preceq C \preceq B \implies F^+(A) \leq F^+(C) \leq F^+(B) \). Yet by construction, we also have \( f(A) = f(B) \neq f(C) \). By the same reasoning as in Theorem 3.3, we have a contradiction. Thus, a DNN\(^+\) cannot solve classification problems that fall into this case.

**case 2** Does NOT exist such a pair \( A \preceq B \). This means for all pairs of points in the two disconnected regions, they don’t follow the ordering defined in Definition 3.3. Thus, there exists two input dimensions, \( i, j \in [n] \), such that for all \( x^0 = (\ldots, x_i^0, \ldots, x_j^0, \ldots) \in R_0 \), and for all \( x^1 = (\ldots, x_i^1, \ldots, x_j^1, \ldots) \in R_1 \), they follow

\[
\begin{align*}
x_i^0 &< t_i < x_i^1 \\
x_j^0 &> t_j > x_j^1 \\
t_1, t_2 &\in \mathbb{R}
\end{align*}
\]

Now, for a point \( G = (\ldots, x_i^F, \ldots, x_j^F, \ldots) \in R_0 \), we always have \( x_i^G < t_i \) and \( x_j^G > t_j \). Further, we can always

\( ^4\)For the choice of \( \epsilon \) under non-linear decision boundary scenario, we assume here we can always find \( \epsilon \) such that it is larger than the linear approximation error. For more details, please see Remark 5.
construct two more points \( D, E \in (K - R_0 - R - 1) \) by

\[
D = (\ldots, x_i^G, \ldots, t_j, \ldots) \\
E = (\ldots, t_i, \ldots, x_j^G, \ldots)
\]

Thus we have \( D \preceq G \preceq E \implies F^+(D) \leq F^+(G) \leq F^+(E) \). However, since we have \( F \in \mathbb{R}_0 \) and yet \( D, E \in (K - R_0 - R - 1) \), we therefore have \( f(D) = f(E) \neq f(G) \). By the same reasoning as in theorem 3.3, we have a contradiction. Thus, a DNN\(^+\) cannot solve classification problems that fall into this case.

Collectively, we proved that a DNN\(^+\) cannot solve a classification problem where there exists a class that is a disconnected space.