5.5 LINEAR SEPARABILITY

Thomas M. Cover

Departments of Electrical Engineering
and Statistics
Stanford University
Stanford, CA 94305

Let \((X_i, \theta_i), i = 1, 2, \ldots, n\) be i.i.d. random pairs, where \(\{ \theta_i \}\) is Bernoulli with parameter \(1/2\), and \(X_i \sim f_\theta(x), x_i \in \mathbb{R}^d\). We say \(\{(X_i, \theta_i)\}_{i=1}^n\) is linearly separable if there exists a vector \(w \in \mathbb{R}^d\) and a real number \(T\) such that

\[
 w^T x_i \geq T, \quad \theta_i = 1
\]

\[
 < T, \quad \theta_i = 0, \quad \text{for} \quad i = 1, 2, \ldots, n.
\]

Let \(P(n, d, f_0, f_1)\) be the associated probability that \(\{(X_i, \theta_i)\}_{i=1}^n\) is linearly separable.

The following results are known.

Theorem 1: Identical distributions [1,2].

\[
 P(n, d, f, f) = 2^{-(n-1)} \sum_{i=0}^{d} \left( \begin{array}{c} n - 1 \\ i \end{array} \right),
\]

for any density \(f(x)\).

Theorem 2: Distributions differing by translation [3].

Let \(f_2(x) = f_1(x + t \nu)\). Then \(P(n, d, f_1, f_2)\) is monotonically increasing in \(t \geq 0\). When \(t = 0\), \(P(n, d, f_1, f_2) = P(n, d, f, f)\), and \(P(n, d, f_1, f_2) \to 1\) as \(t \to \infty\).
Theorem 3. Distribution differing by scale (Krueger, unpublished).

Let \( f_2(x) = \frac{1}{a} f_1(ax), \ a > 0. \) Then \( P(n, d, f_1, f_2) \) is monotonically nondecreasing in \( a \), for \( a \geq 1 \).

All this seems to suggest that different densities lead to an increase in the probability of separability. Hence the following:

Conjecture.

\[
P(n, d, f_1, f_2) \geq \left( \frac{1}{2} \right)^{d-1} \sum_{i=0}^{d} \left( \begin{array}{c} n-1 \\ i \\ \end{array} \right),
\]

for all densities \( f_1(x), f_2(x) \).

REFERENCES

