Introduction to Data-science
Assignment 5 - Additional Content

Alexander Lukjanenkovs (s2545020)
Yulan van Oppen (s2640325)
Emile Ottelé (s3279170)
Amit K. Upadhyay (s3406563)
Jarvin Mutatiina (s3555631)
Parth P. Tiwary (s3576434)

Group 5

October 8, 2018

Unambiguous generalization in homogeneously linearly separable dichotomies

This is a short essay on the article [1] by Thomas M. Cover. In some situations it is necessary for a learning system to be able to separate observations into two disjoint classes. It turns out we require a large number of patterns relative to the number of dimensions for unambiguous generalization. This essentially means that new observations can unambiguously be classified into one of the existing partitions. The paper moreover derives the remarkable result that this classification can be completely characterized by $2d$ patterns, where d is the number of dimensions (or features) under consideration, but we will leave this interesting derivation to the reader.

Suppose we consider a set of N patterns (or: observations) in d dimension (say, with d features). A dichotomy is a partition of a set into two parts. For a set $X = \{x_1, \dots, x_N\}$, a dichotomy $\{X^+, X^-\}$ is linearly separable if there exists a d-dimensional vector of weights w and some real number t satisfying

$$x \cdot w > (<)\ t \quad \text{if} \quad x \in X^+\ (X^-).$$

We say $\{X^+, X^-\}$ is *homogeneously linearly separable* if in the above condition, $t = 0$. The plane $\{\, x \mid x \cdot w = 0 \,\}$ orthogonal to w is called the *separating hyperplane* for $\{X^+, X^-\}$. Suppose we have a d-dimensional vector ϕ of measurement functions $\phi_i : X \to \mathbb{R}$, so that $\phi : X \to \mathbb{R}^d$. Furthermore, a set of N patterns is in general position if any subset of d patterns is linearly independent. An important anchor-point of the paper is the *Function-Counting Theorem*, which states that the number of homogeneously linearly separable dichotomies of N points in general position in $\mathbb{R}^d$ is

$$C(N, d) = 2 \sum_{k=0}^{d-1} \binom{N-1}{k}.$$

We can extend the above-mentioned definitions to the notion of ϕ-separability. A dichotomy $\{X^+, X^-\}$ is ϕ-separable if there exists a d-dimensional vector of weights w such that

$$\phi(x) \cdot w > (<)\ t \quad \text{if} \quad x \in X^+\ (X^-).$$

The plane $\{\, x \mid \phi(x) \cdot w = 0 \,\}$ orthogonal to w is called the *separating ϕ-surface* for $\{X^+, X^-\}$. Moreover, we say X is in ϕ-general position if every subset of cardinality d of $\{\phi(x_1), \dots, \phi(x_N)\}$ is linearly independent.

Assume we randomize the positions of the patterns and the desired dichotomization, and say we have arrived at some set X of N patterns. Along with a dichotomy $\{X^+, X^-\}$ of X, we obtain a training set. Now suppose we observe a new pattern y. The question is, will it belong to X^+ or X^-? We will say the classification of y is *ambiguous relative to a given class of ϕ-surfaces* if two ϕ-surfaces exist such that one assigns y to X^+ and the other assigns it to X^-. Figure illustrates this ambiguity. Here y_1, y_3 are unambiguous, but y_2 is ambiguous with respect to all straight separating lines.

Theorem 6 in [1] shows that y is ambiguous with respect to $C(N, d-1)$ dichotomies of X relative to the class of all ϕ surfaces. Therefore, the probability that y is ambiguous with respect to an arbitrary equiprobable ϕ-separable dichotomy of X turns out to be

$$A(N, d) = \frac{C(N, d-1)}{C(N, d)}.$$

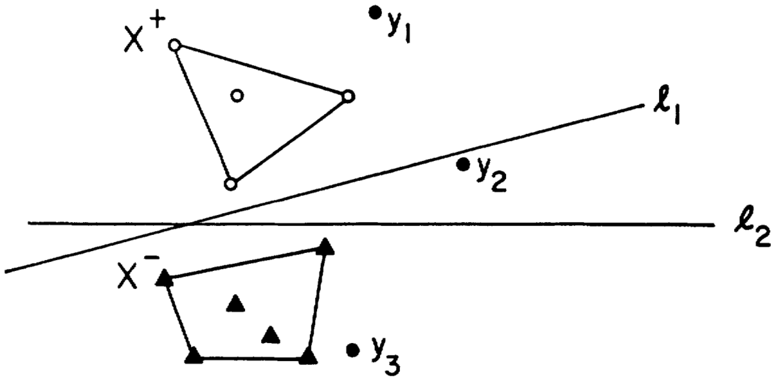

Figure 1: Ambiguous generalization [1]

Even more so, letting $N = \beta d$, where β is such that N is a positive integer, it shows the asymptotic behavior of $A(N, d)$ as $d \to \infty$ to converge to

$$A^*(\beta) = \lim_{d\to\infty} A(N, d) = \begin{cases} 1, & \text{if} \quad \beta \in [0, 2], \\ \frac{1}{\beta - 1}, & \text{if} \quad \beta > 2. \end{cases}$$

This shows a fairly high number of patterns compared to the number of dimensions is required to be able to neglect the probability of ambiguous generalization.

References

[1] Thomas M. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *Electronic Computers, IEEE Transactions on*, EC-14(3):326–334, 1965.