Cover's Function Counting Theorem (1965)

Emin Orhan
eorhan@cns.nyu.edu

December 16, 2014

Problem: Suppose we have p points in $\mathbb{R}^N$. Consider all possible partitions of these p points into two classes. We have 2^p such partitions. How many of these partitions yield linearly separable classes, i.e. where the two classes can be perfectly separated by an $(N-1)$-dimensional hyperplane? The only thing we assume about the points is that they are in general position, which means that any subset of N or fewer points is linearly independent.

Solution: Let's denote the number of linearly separable partitions by $C(p, N)$. We will find an expression for $C(p, N)$ through induction. Imagine first having p points and then adding one more point. Now, considering the linearly separable partitions of the previous p points, there are two possibilities: (1) there is a separating hyperplane for the previous p points passing through the new point, in which case each such linearly separable partition of the previous p points gives rise to two distinct linearly separable partitions when the new point is added as the hyperplane can be shifted infinitesimally to place the new point in either class; (2) there is no separating hyperplane passing through the new point in which case each such linearly separable partition gives rise to only one linearly separable partition when the new point is added, namely the one consistent with the linear separability of all the $p+1$ points. Therefore to find $C(p+1, N)$, we have to count the number of linearly separable partitions in (1) twice and the number of linearly separable partitions in (2) once. This is the same thing as counting the total number of partitions, i.e. $C(p, N)$, and adding the number of linearly separable partitions in (1). But the number of linearly separable partitions in (1) is precisely $C(p, N-1)$, because restricting the separating hyperplane to go through a particular point is the same as eliminating one degree of freedom and thus projecting the p points to a $(N-1)$-dimensional space.

Thus, we have the recursive relation:

$$C(p+1, N) = C(p, N) + C(p, N-1) \tag{1}$$

Iterating this recursion once, we get:

$$C(p+1, N) = C(p-1, N) + 2C(p-1, N-1) + C(p-1, N-2) \tag{2}$$

Iterating once more, we get:

$$C(p+1, N) = C(p-2, N) + 3C(p-2, N-1) + 3C(p-2, N-2) + C(P-2, N-3) \tag{3}$$

It is then easy to check that after $p-1$ iterations we should get:

$$C(p+1, N) = \binom{p}{0} C(1, N) + \binom{p}{1} C(1, N-1) + \ldots + \binom{p}{p} C(1, N-p) \tag{4}$$

We now observe that $C(1, k) = 0$ if $k < 1$. Then, noting that $C(1, k) = 2$ for all k, we get (solid lines in Figure 1):

$$C(p+1, N) = 2 \sum_{i=0}^{N-1} \binom{p}{i} \tag{5}$$

again with the understanding that $\binom{p}{i} = 0$ if $i > p$.

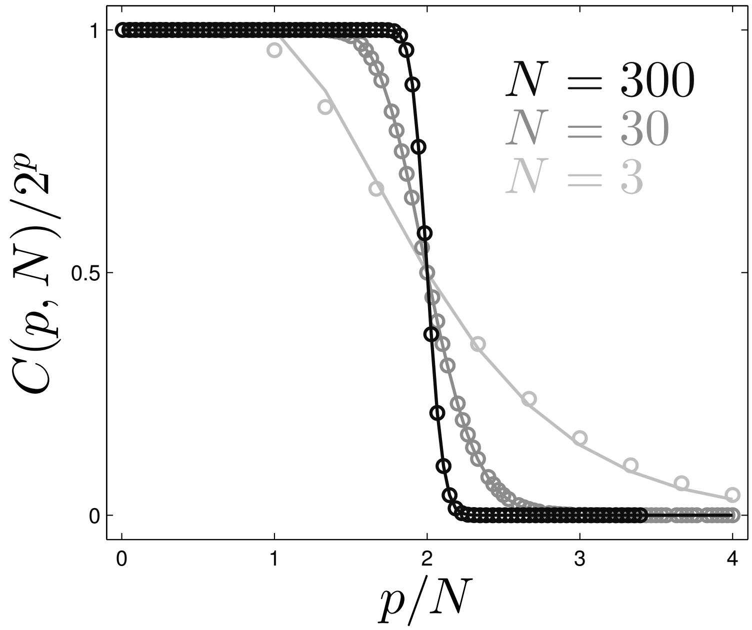

Figure 1: $C(p, N)/2^p$ (fraction of linearly separable partitions) plotted against p/N for three different N values. The solid line uses the exact expression for $C(p, N)$, i.e. Equation 5, the circles use the Gaussian approximation in Equation 11.

For large p, one can use the Gaussian approximation of the binomial coefficients:

$$2^{-m}\binom{m}{n} \approx \mathcal{N}(n; \frac{m}{2}, \frac{m}{4}) \tag{6}$$

where $\mathcal{N}(n; \frac{m}{2}, \frac{m}{4})$ denotes the pdf of a Gaussian distribution with mean $\frac{m}{2}$ and variance $\frac{m}{4}$ evaluated at n. Equation 6 is valid for large m. Using this approximation in Equation 5, we have:

$$\begin{aligned}
\frac{C(p+1,N)}{2^{p+1}} &\approx \sum_{i=0}^{N-1} \mathcal{N}(i; \frac{p}{2}, \frac{p}{4}) && \text{(large } p\text{)} && (7)\\
&\approx \Phi(N-1; \frac{p}{2}, \frac{p}{4}) && \text{(large } p\text{)} && (8)\\
&= \frac{1}{2}\Big[1 + \mathrm{erf}\Big((N-1)\sqrt{\frac{2}{p}} - \sqrt{\frac{p}{2}}\Big)\Big] && \text{(definition of erf)} && (9)\\
&\approx \frac{1}{2}\Big[1 + \mathrm{erf}\Big(N\sqrt{\frac{2}{p}} - \sqrt{\frac{p}{2}}\Big)\Big] && \text{(large } N\text{)} && (10)
\end{aligned}$$

Here, $\Phi(n; m, k)$ denotes the cdf of a Gaussian with mean m and variance k evaluated at n. In the large p limit, we can also replace $p + 1$ with p on the left-hand side to get (circles in Figure 1):

$$\frac{C(p,N)}{2^p} \approx \frac{1}{2}\Big[1 + \mathrm{erf}\Big(N\sqrt{\frac{2}{p}} - \sqrt{\frac{p}{2}}\Big)\Big] \tag{11}$$

References

[1] Hertz J, Krogh A, Palmer RG (1991). *Introduction to the Theory of Neural Computation.* Addison-Wesley.

[2] Cover TM (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. IEEE Trans Electron Comput, 14:326-334.