The Vapnik-Chervonenkis Dimension
Slide Set 12

Prof. Dan A. Simovici

UMB
1 Basic Definitions for Vapnik-Chervonenkis Dimension

2 Growth Functions

3 The Sauer-Shelah Theorem

4 The VCD of Collections of Sets
Trace of a Collection of Sets

Definition

Let \( \mathcal{C} \) be a collection of sets and let \( U \) be a set. The trace of collection \( \mathcal{C} \) on the set \( U \) is the collection

\[
\mathcal{C}_U = \{ U \cap C \mid C \in \mathcal{C} \}.
\]

If the trace of \( \mathcal{C} \) on \( U \), \( \mathcal{C}_U \) equals \( \mathcal{P}(U) \), then we say that \( U \) is shattered by \( \mathcal{C} \).

\( U \) is shattered by \( \mathcal{C} \) if \( \mathcal{C} \) can carve \( \text{any subset of } U \) as an intersection with a set in \( \mathcal{C} \).
Example

Let $U = \{u_1, u_2\}$ and let $\mathcal{C}$ be the collection of sets

$$\mathcal{C} = \{\{u_3\}, \{u_1, u_3\}, \{u_2, u_3\}, \{u_1, u_2, u_3\}\}.$$ 

$\mathcal{C}$ shatters $U$ because we can write:

$$\emptyset = U \cap \{u_3\}$$
$$\{u_1\} = U \cap \{u_1, u_3\}$$
$$\{u_2\} = U \cap \{u_2, u_3\}$$
$$\{u_1, u_2\} = U \cap \{u_1, u_2, u_3\}$$
Definition

The Vapnik-Chervonenkis dimension of the collection $\mathcal{C}$ (called the VC-dimension for brevity) is the largest size of a set $K$ that is shattered by $\mathcal{C}$. This largest size is denoted by $VCD(\mathcal{C})$. 
Example

Note that the previous collection \( \mathcal{C} \) cannot shatter the set \( U' = \{ u_1, u_2, u_3 \} \) because this set has 8 subsets and \( \mathcal{C} \) has just four sets. Thus, it is impossible to express all subsets of \( U' \) as intersections of \( U' \) with some set of \( \mathcal{C} \).

The VCD dimension of the collection \( \mathcal{C} \) is 2.
Note that:

- We have $VCD(\mathcal{C}) = 0$ if and only if $|\mathcal{C}| = 1$.
- If $VCD(\mathcal{C}) = d$, then there exists a set $K$ of size $d$ such that for each subset $L$ of $K$ there exists a set $C \in \mathcal{C}$ such that $L = K \cap C$.
- $\mathcal{C}$ shatters $K$ if and only if the trace of $\mathcal{C}$ on $K$ denoted by $\mathcal{C}_K$ shatters $K$. This allows us to assume without loss of generality that both the sets of the collection $\mathcal{C}$ and a set $K$ shattered by $\mathcal{C}$ are subsets of a set $U$. 
Collections of Sets as Sets of Hypotheses

Let $U$ be a set, $K$ a subset, and let $\mathcal{C}$ be a collection of sets. Each $C \in \mathcal{C}$ defines a hypothesis $h_C : U \rightarrow \{-1, 1\}$ that is a dichotomy, where

$$h_C(u) = \begin{cases} 1 & \text{if } u \in C, \\ -1 & \text{if } u \notin C. \end{cases}$$

$K$ is shattered by $\mathcal{C}$ if and only if for every subset $L$ of $K$ there exists a hypothesis $h_C$ such that the set of positive examples $\{u \in U \mid h_C(u) = 1\}$ equals $L$. 
Finite Collections have Finite VC-Dimension

Let $\mathcal{C}$ be a collection of sets with $VCD(\mathcal{C}) = d$ and let $K$ be a set shattered by $\mathcal{C}$ with $|K| = d$.
Since there exist $2^d$ subsets of $K$, there are at least $2^d$ subsets of $\mathcal{C}$, so

$$2^d \leq |\mathcal{C}|.$$ 

Consequently, $VCD(\mathcal{C}) \leq \log_2 |\mathcal{C}|$. This shows that if $\mathcal{C}$ is finite, then $VCD(\mathcal{C})$ is finite.
The converse is false: there exist infinite collections $\mathcal{C}$ that have a finite VC-dimension.
A Tabular Representation of Collections

If \( U = \{u_1, \ldots, u_n\} \) is a finite set, then the trace of a collection \( \mathcal{C} = \{C_1, \ldots, C_p\} \) of subsets of \( U \) on a subset \( K \) of \( U \) can be presented in an intuitive, tabular form.

Let \( \theta \) be a table containing the rows \( t_1, \ldots, t_p \) and the binary attributes \( u_1, \ldots, u_n \).

Each tuple \( t_k \) corresponds to a set \( C_k \) of \( \mathcal{C} \) and is defined by

\[
t_k[u_i] = \begin{cases} 
1 & \text{if } u_i \in C_k, \\
0 & \text{otherwise},
\end{cases}
\]

for \( 1 \leq i \leq n \). Then, \( \mathcal{C} \) shatters \( K \) if the content of the projection \( r[K] \) consists of \( 2^{|K|} \) distinct rows.
Example

Let $U = \{u_1, u_2, u_3, u_4\}$ and let $\mathcal{C} = \{\{u_2, u_3\}, \{u_1, u_3, u_4\}, \{u_2, u_4\}, \{u_1, u_2\}, \{u_2, u_3, u_4\}\}$ represented by:

\[
\begin{array}{cccc}
  & u_1 & u_2 & u_3 & u_4 \\
0 & 1 & 1 & 0 & \\
1 & 0 & 1 & 1 & \\
0 & 1 & 0 & 1 & \\
1 & 1 & 0 & 0 & \\
0 & 1 & 1 & 1 & \\
\end{array}
\]

The set $K = \{u_1, u_3\}$ is shattered by the collection $\mathcal{C}$ because the projection on $K$ ($(0, 1), (1, 1), (0, 0), (1, 0), (0, 1)$) contains the all four necessary tuples $(0, 1), (1, 1), (0, 0), \text{ and } (1, 0)$.

No subset $K$ of $U$ that contains at least three elements can be shattered by $\mathcal{C}$ because this would require the projection $r[K]$ to contain at least eight tuples. Thus, $VCD(\mathcal{C}) = 2$. 
Observations:

- Every collection of sets shatters the empty set.
- If $C$ shatters a set of size $n$, then it shatters a set of size $p$, where $p \leq n$.

For a collection of sets $C$ and for $m \in \mathbb{N}$, let

$$\Pi_C[m] = \max\{|C_K| \mid |K| = m\}$$

be the largest number of distinct subsets of a set having $m$ elements that can be obtained as intersections of the set with members of $C$.

- We have $\Pi_C[m] \leq 2^m$;
- if $C$ shatters a set of size $m$, then $\Pi_C[m] = 2^m$. 
A Vapnik-Chervonenkis class (or a VC class) is a collection \( \mathcal{C} \) of sets such that \( VCD(\mathcal{C}) \) is finite.
Example

Let $\mathbb{R}$ be the set of real numbers and let $I$ be the collection of sets $\{(-\infty, t) \mid t \in \mathbb{R}\}$.

We claim that any singleton is shattered by $I$. Indeed, if $S = \{x\}$ is a singleton, then $P(\{x\}) = \{\emptyset, \{x\}\}$. Thus, if $t \geq x$, we have $(-\infty, t) \cap S = \{x\}$; also, if $t < x$, we have $(-\infty, t) \cap S = \emptyset$, so $I_S = P(S)$.

There is no set $S$ with $|S| = 2$ that can be shattered by $I$. Indeed, suppose that $S = \{x, y\}$, where $x < y$. Then, any member of $I$ that contains $y$ includes the entire set $S$, so $I_S = \{\emptyset, \{x\}, \{x, y\}\} \neq P(S)$. This shows that $I$ is a VC class and $\text{VCD}(I) = 1$. 
Consider the collection $I = \{[a, b] \mid a, b \in \mathbb{R}, a \leq b\}$ of closed intervals. We claim that $VCD(I) = 2$. To justify this claim, we need to show that there exists a set $S = \{x, y\}$ such that $I_S = \mathcal{P}(S)$ and no three-element set can be shattered by $I$.

For the first part of the statement, consider the intersections

$$
\begin{align*}
[u, v] \cap S &= \emptyset, \text{ where } v < x, \\
[x - \epsilon, \frac{x+y}{2}] \cap S &= \{x\}, \\
[\frac{x+y}{2}, y] \cap S &= \{y\}, \\
[x - \epsilon, y + \epsilon] \cap S &= \{x, y\},
\end{align*}
$$

which show that $I_S = \mathcal{P}(S)$.

For the second part of the statement, let $T = \{x, y, z\}$ be a set that contains three elements. Any interval that contains $x$ and $z$ also contains $y$, so it is impossible to obtain the set $\{x, z\}$ as an intersection between an interval in $I$ and the set $T$. 


An Example

Let $\mathcal{H}$ be the collection of closed half-planes in $\mathbb{R}^2$ of the form

$$\{x = (x_1, x_2) \in \mathbb{R}^2 \mid ax_1 + bx_2 - c \geq 0, a \neq 0 \text{ or } b \neq 0\}.$$

We claim that $VCD(\mathcal{H}) = 3$.

Let $P, Q, R$ be three non-colinear points. Each line is marked with the sets it defines; thus, it is clear that the family of half-planes shatters the set $\{P, Q, R\}$, so $VCD(\mathcal{H})$ is at least 3.
Example (cont’d)

To complete the justification of the claim we need to show that no set that contains at least four points can be shattered by $\mathcal{H}$.

Let $\{P, Q, R, S\}$ be a set that contains four points such that no three points of this set are collinear. If $S$ is located inside the triangle $P, Q, R$, then every half-plane that contains $P, Q, R$ also contains $S$, so it is impossible to separate the subset $\{P, Q, R\}$. Thus, we may assume that no point is inside the triangle formed by the remaining three points.

Any half-plane that contains two diagonally opposite points, for example, $P$ and $R$, contains either $Q$ or $S$, which shows that it is impossible to separate the set $\{P, R\}$. Thus, no set that contains four points may be shattered by $\mathcal{H}$, so $VCD(\mathcal{H}) = 3$. 
CLAIM: the VCD of an arbitrary family of hyperplanes in $\mathbb{R}^d$ is $d + 1$. Consider the set of $d + 1$ points $\{x_0, x_1, \ldots, x_d\}$ defined as

$$x_0 = 0_d, x_i = e_1 \text{ for } 1 \leq i \leq d.$$ 

Let $y_0, y_1, \ldots, y_d \in \{-1, 1\}$ and let $w \in \mathbb{R}^d$ be the vector whose $i^{\text{th}}$ coordinate is $y_i$. We have $w'x = y_i$ for $1 \leq i \leq d$. Therefore,

$$\text{sign } \left( w'x_i + \frac{y_0}{2} \right) = \text{sign } \left( y_i + \frac{y_0}{2} \right) = y_i.$$ 

Thus, points $x_i$ for which $y_i = 1$ are on the positive side of the hyperplane $y'x = 0$; the ones for which $y_i = -1$ are on the opposite side, so any family of $d + 1$ points in $\mathbb{R}^d$ can be shattered by hyperplanes.
Also we need to show that no set of $d + 2$ points can be shattered by hyperplanes. For this we need the notion of convex set and the notion of convex hull.
Let \( \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \). The \emph{closed segment} determined by \( \mathbf{x} \) and \( \mathbf{y} \) is the set 

\[
[x, y] = \{(1 - a)x + ay \mid 0 \leq a \leq 1\}.
\]

**Definition**

A subset \( C \) of \( \mathbb{R}^n \) is \emph{convex} if, for all \( \mathbf{x}, \mathbf{y} \in C \) we have \([x, y] \subseteq C\).
Convex Set (a) vs. a Non-convex Set (b)
Example

The convex subsets of $\mathbb{R}$ are the intervals of $\mathbb{R}$. Regular polygons are convex subsets of $\mathbb{R}^2$. An open sphere $B(x_0, r)$ or a closed sphere $B[x_0, r]$ in $\mathbb{R}^n$ is convex.
Definition

Let $U$ be a subset of $\mathbb{R}^n$. A convex combination of $U$ is a vector of the form $a_1x_1 + \cdots + a_kx_k$, where $x_1, \ldots, x_k \in U$, $a_i \geq 0$ for $1 \leq i \leq k$, and $a_1 + \cdots + a_k = 1$. 
Theorem

The intersection of any collection of convex sets in $\mathbb{R}^n$ is a convex set.

Proof.

Let $\mathcal{C} = \{C_i \mid i \in I\}$ be a collection of convex sets and let $C = \bigcap \mathcal{C}$. Suppose that $x_1, \ldots, x_k \in C$, $a_i \geq 0$ for $1 \leq i \leq k$, and $a_1 + \cdots + a_k = 1$. Since $x_1, \ldots, x_k \in C_i$, it follows that $a_1x_1 + \cdots + a_kx_k \in C_i$ for every $i \in I$. Thus, $a_1x_1 + \cdots + a_kx_k \in C$, which proves the convexity of $C$. □
Definition

The **convex hull** (or the **convex closure**) of a subset $U$ of $\mathbb{R}^n$ is the intersection of all convex sets that contain $U$, that is, the **smallest convex set** that contains $U$.

The convex null of $U$ is denoted by $\mathbf{K}_{\text{conv}}(U)$.
Theorem

Let $S$ be a subset of $\mathbb{R}^n$. The convex hull $K_{conv}(S)$ consists of the set of all convex combinations of elements of $S$, that is,

$$K_{conv}(S) = \{ a_1x_1 + \cdots + a_mx_m, x_1, \ldots, x_m \in S \mid a_1, \ldots, a_m \geq 0 \text{ and } \sum_{i=1}^{m} a_i = 1 \}.$$
Proof

Note that $S \subseteq \mathcal{K}_{\text{conv}}(S)$ because $x \in S$ implies $1x = x \in \mathcal{K}_{\text{conv}}(S)$. The set $\mathcal{K}_{\text{conv}}(S)$ is convex. Indeed, let

$$
u = a_1x_1 + \cdots + a_mx_m \in \mathcal{K}_{\text{conv}}(S),$$

$$v = b_1x_1 + \cdots + b_mx_m \in \mathcal{K}_{\text{conv}}(S),$$

$$a_1, \ldots, a_m \geq 0 \quad \text{and} \quad \sum_{i=1}^{m} a_i = 1,$$

$$b_1, \ldots, b_m \geq 0 \quad \text{and} \quad \sum_{i=1}^{m} b_i = 1,$$

where we assume, without loss of generality, that the two convex combinations involve the same number of terms.
Let $c \in [0, 1]$ and let $z = cu + (1 - c)v$.
Since
\[ z = \sum_{i=1}^{m} (ca_i + (1 - c)b_i)x_i \]
and $\sum_{i=1}^{m} (ca_i + (1 - c)b_i) = c \sum_{i=1}^{m} a_i + (1 - c) \sum_{i=1}^{m} b_i = 1$, it follows that $z \in K_{\text{conv}}(S)$, so $K_{\text{conv}}(S)$ is convex.
Proof continued

Every convex set $T$ that contains $S$ will contain $K_{\text{conv}}(S)$, hence $K_{\text{conv}}(S)$ is the smallest convex set that contains $S$. 
Example

A two-dimensional simplex is defined starting from three points $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ in $\mathbb{R}^2$ such that none of these points is collinear with the others two. Thus, the two-dimensional simplex generated by $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ is the full triangle determined by $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$. 

\[ x_1 \quad x_2 \quad x_3 \]
Let $S$ be the $n$-dimensional simplex generated by the points $x_1, \ldots, x_{n+1}$ in $\mathbb{R}^n$ and let $x \in S$. If $x \in S$, then $x$ is a convex combination of $x_1, \ldots, x_n, x_{n+1}$. In other words, there exist $a_1, \ldots, a_n, a_{n+1}$ such that $a_1, \ldots, a_n, a_{n+1} \in (0, 1)$, $\sum_{i=1}^{n+1} a_i = 1$, and

$$x = a_1 x_1 + \cdots + a_n x_n + a_{n+1} x_{n+1}.$$
(Radon’s Theorem) Any set $X = \{x_1, \ldots, x_{d+2}\}$ of $d + 2$ points in $\mathbb{R}^d$ can be partitioned into two sets $X_1$ and $X_2$ such that the convex hulls of $X_1$ and $X_2$ intersect.
Proof

Consider the following system with $d + 1$ linear equations and $d + 2$ variables $\alpha_1, \alpha_2, \ldots, \alpha_{d+2}$:

\[
\sum_{i=1}^{d+2} \alpha_i x_i = 0_d, \quad (d \text{ scalar equations})
\]
\[
\sum_{i=1}^{d+2} \alpha_i = 0.
\]

Since the number of variables $d + 2$ is larger than the number of equations $d + 1$, the system has a non-trivial solution $\beta_1, \ldots, \beta_{d+2}$. Since $\sum_{i=1}^{d+2} \beta_i = 0$ both sets

\[
l_1 = \{i | 1 \leq i \leq d + 2, \beta_i > 0\}, \quad l_2 = \{i | 1 \leq i \leq d + 2, \beta_i < 0\}
\]

are non-empty sets and disjoint sets, and

\[
X_1 = \{x_i \mid i \in l_1\}, \quad X_2 = \{x_i \mid i \in l_2\},
\]

form a partition of $X$. 
Proof (cont’d)

Define $\beta = \sum_{i \in I_1} \beta_i$. Since $\sum_{i \in I_1} \beta_i = -\sum_{i \in I_2} \beta_i$, we have

$$\sum_{i \in I_1} \frac{\beta_i}{\beta} \mathbf{x}_i = \sum_{i \in I_2} \frac{-\beta_i}{\beta} \mathbf{x}_i.$$

Also,

$$\sum_{i \in I_1} \frac{\beta_i}{\beta} = \sum_{i \in I_2} \frac{-\beta_i}{\beta} = 1,$$

$\frac{\beta_i}{\beta} \geq 0$ for $i \in I_1$ and $\frac{-\beta_i}{\beta} \geq 0$ for $i \in I_2$. This implies that

$$\sum_{i \in I_1} \frac{\beta_i}{\beta} \mathbf{x}_i$$

belongs both to the convex hulls of $X_1$ and $X_2$. 
Let $X$ be a set of $d + 2$ points in $\mathbb{R}^d$. By Radon’s Theorem it can be partitioned into $X_1$ and $X_2$ such that the two convex hulls intersect. When two sets are separated by a hyperplane, their convex hulls are also separated by the hyperplane. Thus, $X_1$ and $X_2$ cannot be separated by a hyperplane and $X$ is not shattered.
Example

Let $\mathcal{R}$ be the set of rectangles whose sides are parallel with the axes $x$ and $y$. There is a set $S$ with $|S| = 4$ that is shattered by $\mathcal{R}$. Let $S$ be a set of four points in $\mathbb{R}^2$ that contains a unique “northernmost point” $P_n$, a unique “southernmost point” $P_s$, a unique “easternmost point” $P_e$, and a unique “westernmost point” $P_w$. If $L \subseteq S$ and $L \neq \emptyset$, let $R_L$ be the smallest rectangle that contains $L$. For example, we show the rectangle $R_L$ for the set $\{P_n, P_s, P_e\}$.
This collection cannot shatter a set of points that contains at least five points. Indeed, let $S$ be such that $|S| \geq 5$. If the set contains more than one “northernmost” point, then we select exactly one to be $P_n$. Then, the rectangle that contains the set $K = \{ P_n, P_e, P_s, P_w \}$ contains the entire set $S$, which shows the impossibility of separating $S$. 
The Class of All Convex Polygons

Example

Consider the system of all convex polygons in the plane. For any positive integer $m$, place $m$ points on the unit circle. Any subset of the points are the vertices of a convex polygon. Clearly that polygon will not contain any of the points not in the subset. This shows that we can shatter arbitrarily large sets, so the VC-dimension of the class of all convex polygons is infinite.
The Case of Convex Polygons with $d$ Vertices

Example

Consider the class of \textit{convex polygons that have no more than $d$ vertices in $\mathbb{R}^2$} and place $2d + 1$ points on a circle.

- Label a subset of these points as positive, and the remaining points as negative. Since we have an odd number of points there exists a majority in one of the classes (positive or negative).
- If the negative point are in majority, there are at most $d$ positive points; these are contained by the convex polygon formed by joining the positive points.
- If the positive are in majority, consider the polygon formed by the tangents of the negative points.
Negative Points in the Majority

positive examples

negative examples
Positive Points in the Majority

- positive examples
- negative examples
Since a set with $2d + 1$ points can be shattered, the VC dimension of the set of convex polygons with at most $d$ vertices is at least $2d + 1$.

If all labeled points are located on a circle then it is impossible for a point to be in the convex closure of a subset of the remaining points. Thus, placing the points on a circle maximizes the number of sets required to shatter the set, so the VC-dimension is indeed $2d + 1$. 
Definition

Let $H$ be a set of hypotheses and let $(x_1, \ldots, x_m)$ be a sequence of examples of length $m$. A hypothesis $h \in H$ induces a classification

$$(h(x_1), \ldots, h(x_m))$$

of the components of this sequence. Note that the number of ways in which $h$ can classify the members of the sequence $(x_1, \ldots, x_m)$ is $|\{h(x_1), \ldots, h(x_m)\}|$.

The growth function of $H$ is the function $\Pi_H : \mathbb{N} \rightarrow \mathbb{N}$ gives the number of ways a sequence of examples of length $m$ can be classified by a hypothesis in $H$:

$$\Pi_H(m) = \max_{(x_1, \ldots, x_m) \in \mathcal{X}^m} \{||\{(h(x_1), \ldots, h(x_m))\}| \mid h \in H\}$$
Definition

A dichotomy is a hypothesis \( h : \mathcal{X} \rightarrow \{-1, 1\} \).

If \( H \) consists of dichotomies, then \((x_1, \ldots, x_m)\) can be classified in at most \(2^m\) ways.
A Preliminary Result

**Theorem**

Let $S = \{s_1, \ldots, s_n\}$ be a set and let $\mathcal{C}$ be a collection of subsets of $S$, $\mathcal{C} \subseteq \mathcal{P}(S)$.

Let $SH(\mathcal{C})$ be the family of subsets of $S$ that are shattered by $\mathcal{C}$.

Then, we have $|SH(\mathcal{C})| \geq |\mathcal{C}|$. 
Proof

The argument is by induction on $|\mathcal{C}|$.
Consider the subfamilies

\[
\mathcal{C}_0 = \{ U \in \mathcal{C} \mid s_1 \not\in U \} \\
\mathcal{C}_1 = \{ U \in \mathcal{C} \mid s_1 \in U \}
\]

By the inductive hypothesis, $|\text{SH}(\mathcal{C}_0)| \geq |\mathcal{C}_0|$, that is, $\mathcal{C}_0$ shatters at least as many subsets of $S' = \{s_2, s_3, \ldots, s_n\}$ as $|\mathcal{C}_0|$.
Next, consider the family

\[
\mathcal{C}_1' = \{ U - \{s_1\} \mid U \in \mathcal{C}_1 \}.
\]

Note that:

- The families $\mathcal{C}_0$ and $\mathcal{C}_1$ of subsets of $S$ are disjoint and $|\mathcal{C}| = |\mathcal{C}_0| + |\mathcal{C}_1|$.
- $\mathcal{C}_0$ and $\mathcal{C}_1'$ are families of subsets of $S'$ and $|\mathcal{C}_1'| = |\mathcal{C}_1'|$. 
By induction, $C'_1$ shatters at least as many subsets of $S' = \{s_2, s_3, \ldots, s_n\}$ as its cardinality, that is, $|\text{SH}(C'_1)| \geq |C'_1|$. The number of subsets of $S'$ shattered by $C_0$ and $C'_1$ sum up to at least $|C_0| + |C'_1| = |C_0| + |C_1| = |C|$, and every subset of $S'$ shattered by $C'_1$ is shattered by $C_1 \subseteq C$. Note that there may be subsets $V$ of $S'$ shattered by both $C_0$ and $C'_1$. In this case both $V$ and $V \cup \{s_1\}$ are shattered by $C$. 

For $n, k \in \mathbb{N}$ and $0 \leq k \leq n$ define the number $\binom{n}{\leq k}$ as:

$$\binom{n}{\leq k} = \sum_{i=0}^{k} \binom{n}{i}.$$  

Clearly, $\binom{n}{\leq 0} = 1$ and $\binom{n}{\leq n} = 2^n$.

Observe that if $\mathcal{P}_k(S)$ is the collection of subsets of $S$ that contain $k$ or fewer elements, then for $|S| = n$,

$$|\mathcal{P}_k(S)| = \binom{n}{\leq k}.$$
(Sauer-Shelah Theorem) Let $S$ be a set with $|S| = n$ and let $\mathcal{C}$ be a collection of subsets of $S$ such that

$$|\mathcal{C}| > \binom{n}{\leq k}$$

Then, there exists a subset $T$ of $S$ having at least $k + 1$ elements such that $\mathcal{C}$ shatters $T$. 
Proof

Let $|SH(C)|$ be the number of sets shattered by $C$. We have $|SH(C)| \geq |C|$ by the previous theorem. The inequality of the theorem means that $|C| > |P_k(S)|$, hence $|SH(C)| > |P_k(S)|$. Therefore, there exists a subset $T$ of $S$ with at least $k + 1$ elements that is shattered by $C$. 
Theorem

Let $\phi : \mathbb{N}^2 \rightarrow \mathbb{N}$ be the function defined by

$$\phi(d, m) = \begin{cases} 
1 & \text{if } m = 0 \text{ or } d = 0 \\
\phi(d, m - 1) + \phi(d - 1, m - 1), & \text{otherwise.}
\end{cases}$$

We have

$$\phi(d, m) = \binom{m}{\leq d}$$

for $d, m \in \mathbb{N}$. 

Proof

The argument is by strong induction on $s = d + m$. The base case, $s = 0$, implies $m = 0$ and $d = 0$, and the equality is immediate.
Suppose that the equality holds for \( \phi(d', m') \), where \( d' + m' < d + m \). We have:

\[
\phi(d, m) = \phi(d, m - 1) + \phi(d - 1, m - 1)
\]
(by definition)
\[
= \sum_{i=0}^{d} \binom{m-1}{i} + \sum_{i=0}^{d-1} \binom{m-1}{i}
\]
(by inductive hypothesis)
\[
= \sum_{i=0}^{d} \binom{m-1}{i} + \sum_{i=1}^{d} \binom{m-1}{i-1}
\]
(by changing the summation index in the second sum)
\[
= \sum_{i=0}^{d} \binom{m-1}{i} + \sum_{i=0}^{d} \binom{m-1}{i-1}
\]
(because \( \binom{m-1}{-1} = 0 \))
\[
= \sum_{i=0}^{d} \left( \binom{m-1}{i} + \binom{m-1}{i-1} \right)
\]
\[
= \sum_{i=0}^{d} \binom{m}{i} = \binom{m}{\leq d},
\]

which gives the desired conclusion.
Another Inequality

Suppose that \( VCD(C) = d \) and \(|S| = n\). Then \( SH(C) \subseteq \mathcal{P}_d(S) \), hence

\[
|C| \leq |SH(C)| \leq \sum_{i=1}^{d} \binom{n}{i} = \binom{n}{\leq d}.
\]

Together with the previous inequality we obtain:

\[
2^d \leq |C| \leq \binom{n}{\leq d} = \phi(n, d).
\]
Lemma

For $d \in \mathbb{N}$ and $d \geq 2$ we have

$$2^{d-1} \leq \frac{d^d}{d!}.$$ 

Proof: The argument is by induction on $d$. In the basis step, $d = 2$ both members are equal to 2.
Suppose the inequality holds for $d$. We have

$$
\frac{(d + 1)^{d+1}}{(d + 1)!} = \frac{(d + 1)^d}{d!} = \frac{d^d}{d!} \cdot \frac{(d + 1)^d}{d^d}
$$

$$
= \frac{d^d}{d!} \cdot \left(1 + \frac{1}{d}\right)^d \geq 2^d \cdot \left(1 + \frac{1}{d}\right)^d \geq 2^{d+1}
$$

(by inductive hypothesis)

because

$$
\left(1 + \frac{1}{d}\right)^d \geq 1 + d \frac{1}{d} = 2.
$$

This concludes the proof of the inequality.
Lemma

We have $\phi(d, m) \leq 2\frac{m^d}{d!}$ for every $m \geq d$ and $d \geq 1$.

**Proof:** The argument is by induction on $d$ and $n$. If $d = 1$, then $\phi(1, m) = m + 1 \leq 2m$ for $m \geq 1$, so the inequality holds for every $m \geq 1$, when $d = 1$. 
Proof (cont’d)

If \( m = d \geq 2 \), then \( \phi(d, m) = \phi(d, d) = 2^d \) and the desired inequality follows immediately from a previous Lemma.

Suppose that the inequality holds for \( m > d \geq 1 \). We have

\[
\phi(d, m + 1) = \phi(d, m) + \phi(d - 1, m)
\]

(by the definition of \( \phi \))

\[
\leq 2 \frac{m^d}{d!} + 2 \frac{m^{d-1}}{(d-1)!}
\]

(by inductive hypothesis)

\[
= 2 \frac{m^{d-1}}{(d-1)!} \left( 1 + \frac{m}{d} \right).
\]
Proof (cont’d)

It is easy to see that the inequality

\[ 2 \frac{m^{d-1}}{(d-1)!} \left(1 + \frac{m}{d}\right) \leq \frac{2(m+1)^d}{d!} \]

is equivalent to

\[ \frac{d}{m} + 1 \leq \left(1 + \frac{1}{m}\right)^d \]

and, therefore, is valid. This yields immediately the inequality of the lemma.
The Asymptotic Behavior of the Function $\phi$

**Theorem**

The function $\phi$ satisfies the inequality:

$$\phi(d, m) < \left( \frac{em}{d} \right)^d$$

for every $m \geq d$ and $d \geq 1$.

**Proof:** By a previous Lemma, $\phi(d, m) \leq 2 \frac{m^d}{d!}$. Therefore, we need to show only that

$$2 \left( \frac{d}{e} \right)^d < d!.$$
Proof (cont’d)

The argument is by induction on \( d \geq 1 \). The basis case, \( d = 1 \) is immediate. Suppose that \( 2 \left( \frac{d}{e} \right)^{d} < d! \). We have

\[
2 \left( \frac{d + 1}{e} \right)^{d+1} = 2 \left( \frac{d}{e} \right)^{d} \left( \frac{d + 1}{d} \right)^{d} \frac{d + 1}{e}
\]

\[
= \left( 1 + \frac{1}{d} \right)^{d} \frac{1}{e} \cdot 2 \left( \frac{d}{e} \right)^{d} (d + 1) < 2 \left( \frac{d}{e} \right)^{d} (d + 1),
\]

because

\[
\left( 1 + \frac{1}{d} \right)^{d} < e.
\]

The last inequality holds because the sequence \( \left( (1 + \frac{1}{d})^{d} \right)_{d \in \mathbb{N}} \) is an increasing sequence whose limit is \( e \). Since \( 2 \left( \frac{d+1}{e} \right)^{d+1} < 2 \left( \frac{d}{e} \right)^{d} (d + 1) \), by inductive hypothesis we obtain:

\[
2 \left( \frac{d + 1}{e} \right)^{d+1} < (d + 1)!.
\]

This proves the inequality of the theorem.
Corollary

If $m$ is sufficiently large we have $\phi(d, m) = O(m^d)$.

The statement is a direct consequence of the previous theorem.
Denote by $\oplus$ the symmetric difference of two sets.

**Theorem**

Let $\mathcal{C}$ a family of sets and $C_0 \in \mathcal{C}$. Define the family $\Delta_{C_0}$ as

$$\Delta_{C_0}(\mathcal{C}) = \{ T \mid T = C_0 \oplus C \text{ where } C \in \mathcal{C} \}.$$

We have $VCD(\mathcal{C}) = VCD(\Delta_{C_0}(\mathcal{C}))$. 
Proof

Let $S$ be a set, $S = C_S$ and $S_0 = (\Delta_{C_0}(C))_S$. Define $\psi : S \rightarrow S_0$ as $\psi(S \cap C) = S \cap (C_0 \oplus C)$. We claim that $\psi$ is a bijection.

If $\psi(S \cap C) = \psi(S \cap C')$ for $C, C' \in C$, then $S \cap (C_0 \oplus C) = S \cap (C_0 \oplus C')$. Therefore,

$$(S \cap C_0) \oplus (S \cap C) = (S \cap C_0) \oplus (S \cap C'),$$

which implies $S \cap C = S \cap C'$, so $\psi$ is injective.

On other hand, if $U \in S_0$ we have $U = S \cap (C_0 \oplus C)$, so $U = \psi(S \cap C)$, hence $\psi$ is a surjection. Thus, $S$ and $S_0$ have the same number of sets, which implies that a set $S$ is shattered by $C$ if and only if it is shattered by $\Delta_{C_0}(C)$. 


Let $u : B_2^k \rightarrow B_2$ be a Boolean function of $k$ arguments and let $C_1, \ldots, C_k$ be $k$ subsets of a set $U$. Define the set $u(C_1, \ldots, C_k)$ as the subset $C$ of $U$ whose indicator function is $I_C = u(I_{C_1}, \ldots, I_{C_k})$.

Example

If $u : B_2^2 \rightarrow B_2$ is the Boolean function $u(a_1, a_2) = a_1 \lor a_2$, then $u(C_1, C_2)$ is $C_1 \cup C_2$; similarly, if $u(x_1, x_2) = x_1 \oplus x_2$, then $u(C_1, C_2)$ is the symmetric difference $C_1 \oplus C_2$ for every $C_1, C_2 \in \mathcal{P}(U)$. 
Let $u : B_2^k \rightarrow B_2$ and $C_1, \ldots, C_k$ are $k$ family of subsets of $U$, the family of sets $u(C_1, \ldots, C_k)$ is

$$u(C_1, \ldots, C_k) = \{u(C_1, \ldots, C_k) \mid C_1 \in C_1, \ldots, C_k \in C_k\}.$$ 

**Theorem**

Let $\alpha(k)$ be the least integer $a$ such that $\frac{a}{\log(ea)} > k$. If $C_1, \ldots, C_k$ are $k$ collections of subsets of the set $U$ such that $d = \max\{VCD(C_i) \mid 1 \leq i \leq k\}$ and $u : B_2^2 \rightarrow B_2$ is a Boolean function, then

$$VCD(u(C_1, \ldots, C_k)) \leq \alpha(k) \cdot d.$$
Proof

Let $S$ be a subset of $U$ that consists of $m$ elements. The collection $(C_i)_S$ is not larger than $\phi(d, m)$. For a set in the collection $W \in u(C_1, \ldots, C_k)_S$ we can write $W = S \cap u(C_1, \ldots, C_k)$, or, equivalently,

$1_W = 1_S \cdot u(1_{C_1}, \ldots, 1_{C_k})$.

There exists a Boolean function $g_S$ such that

$1_S \cdot u(1_{C_1}, \ldots, 1_{C_k}) = g_S(1_S \cdot 1_{C_1}, \ldots, 1_S \cdot 1_{C_k}) = g_S(1_S \cap C_1, \ldots, 1_S \cap C_k)$.

Since there are at most $\phi(d, m)$ distinct sets of the form $S \cap C_i$ for every $i$, $1 \leq i \leq k$, it follows that there are at most $(\phi(d, m))^k$ distinct sets $W$, hence $u(C_1, \ldots, C_k)[m] \leq (\phi(d, m))^k$. 
By a previous theorem,

\[ u(C_1, \ldots, C_k)[m] \leq \left( \frac{em}{d} \right)^{kd}. \]

We observed that if \( \Pi_C[m] < 2^m \), then \( VCD(C) < m \). Therefore, to limit the Vapnik-Chervonenkis dimension of the collection \( u(C_1, \ldots, C_k) \) it suffices to require that \( \left( \frac{em}{d} \right)^{kd} < 2^m \).

Let \( a = \frac{m}{d} \). The last inequality can be written as \( (ea)^{kd} < 2^{ad} \); equivalently, we have \( (ea)^k < 2^a \), which yields \( k < \frac{a}{\log(ea)} \). If \( \alpha(k) \) is the least integer \( a \) such that \( k < \frac{a}{\log(ea)} \), then \( m \leq \alpha(k)d \), which gives our conclusion.
Example

If $k = 2$, the least integer $a$ such that $\frac{a}{\log(ea)} > 2$ is $k = 10$, as it can be seen by graphing this function; thus, if $\mathcal{C}_1, \mathcal{C}_2$ are two collection of concepts with $VCD(\mathcal{C}_1) = VCD(\mathcal{C}_2) = d$, the Vapnik-Chervonenkis dimension of the collections $\mathcal{C}_1 \lor \mathcal{C}_2$ or $\mathcal{C}_1 \land \mathcal{C}_2$ is not larger than $10d$. 
Lemma

Let $S, T$ be two sets and let $f : S \rightarrow T$ be a function. If $\mathcal{D}$ is a collection of subsets of $T$, $U$ is a finite subset of $S$ and $\mathcal{C} = f^{-1}(\mathcal{D})$ is the collection $\{ f^{-1}(D) \mid D \in \mathcal{D} \}$, then $|\mathcal{C}_U| \leq |\mathcal{D}_{f(U)}|$.

Proof: Let $V = f(U)$ and denote $f \upharpoonright_U$ by $g$. For $D, D' \in \mathcal{D}$ we have

$$
(U \cap f^{-1}(D)) \oplus (U \cap f^{-1}(D'))
= U \cap (f^{-1}(D) \oplus f^{-1}(D')) = U \cap (f^{-1}(D \oplus D'))
= g^{-1}(V \cap (D \oplus D')) = g^{-1}(V \cap D) \oplus g^{-1}(V \oplus D').
$$

Thus, $C = U \cap f^{-1}(D)$ and $C' = U \cap f^{-1}(D')$ are two distinct members of $\mathcal{C}_U$, then $V \cap D$ and $V \cap D'$ are two distinct members of $\mathcal{D}_{f(U)}$. This implies $|\mathcal{C}_U| \leq |\mathcal{D}_{f(U)}|$. 
Theorem

Let $S, T$ be two sets and let $f : S \rightarrow T$ be a function. If $\mathcal{D}$ is a collection of subsets of $T$ and $\mathcal{C} = f^{-1}(\mathcal{D})$ is the collection $\{f^{-1}(D) \mid D \in \mathcal{D}\}$, then $VCD(\mathcal{C}) \leq VCD(\mathcal{D})$. Moreover, if $f$ is a surjection, then $VCD(\mathcal{C}) = VCD(\mathcal{D})$. 
Proof

Suppose that $C$ shatters an $n$-element subset $K = \{x_1, \ldots, x_n\}$ of $S$, so $|C_K| = 2^n$ By a previous Lemma we have $|C_K| \leq |D_{f(U)}|$, so $|D_{f(U)}| \geq 2^n$, which implies $|f(U)| = n$ and $|D_{f(U)}| = 2^n$, because $f(U)$ cannot have more than $n$ elements. Thus, $D$ shatters $f(U)$, so $VCD(C) \leq VCD(D)$.

Suppose now that $f$ is surjective and $H = \{t_1, \ldots, t_m\}$ is an $m$ element set that is shattered by $D$. Consider the set $L = \{u_1, \ldots, u_m\}$ such that $u_i \in f^{-1}(t_i)$ for $1 \leq i \leq m$. Let $U$ be a subset of $L$. Since $H$ is shattered by $D$, there is a set $D \in D$ such that $f(U) = H \cap D$, which implies $U = L \cap f^{-1}(D)$. Thus, $L$ is shattered by $C$ and this means that $VCD(C) = VCD(D)$. 
Definition

The density of $\mathcal{C}$ is the number

$$\text{dens}(\mathcal{C}) = \inf\{s \in \mathbb{R}_{>0} \mid \prod_{\mathcal{C}}[m] \leq c \cdot m^s \text{ for every } m \in \mathbb{N}\},$$

for some positive constant $c$. 
Theorem

Let $S, T$ be two sets and let $f : S \rightarrow T$ be a function. If $\mathcal{D}$ is a collection of subsets of $T$ and $\mathcal{C} = f^{-1}(\mathcal{D})$ is the collection \{ $f^{-1}(D) \mid D \in \mathcal{D}$ \}, then $\text{dens}(\mathcal{C}) \leq \text{dens}(\mathcal{D})$. Moreover, if $f$ is a surjection, then $\text{dens}(\mathcal{C}) = \text{dens}(\mathcal{D})$.

Proof: Let $L$ be a subset of $S$ such that $|L| = m$. Then, $|\mathcal{C}_L| \leq |\mathcal{D}_{f(L)}|$. In general, we have $|f(L)| \leq m$, so $|\mathcal{D}_{f(L)}| \leq \mathcal{D}[m] \leq cm^s$. Therefore, we have $|\mathcal{C}_L| \leq |\mathcal{D}_{f(L)}| \leq \mathcal{D}[m] \leq cm^s$, which implies $\text{dens}(\mathcal{C}) \leq \text{dens}(\mathcal{D})$. If $f$ is a surjection, then, for every finite subset $M$ of $T$ such that $|M| = m$ there is a subset $L$ of $S$ such that $|L| = |M|$ and $f(L) = M$. Therefore, $\mathcal{D}[m] \leq \Pi_{\mathcal{C}}[m]$ and this implies $\text{dens}(\mathcal{C}) = \text{dens}(\mathcal{D})$. 
If \( C, D \) are two collections of sets such that \( C \subseteq D \), then
\[
VCD(C) \leq VCD(D) \quad \text{and} \quad \text{dens}(C) \leq \text{dens}(D).
\]

**Theorem**

Let \( C \) be a collection of subsets of a set \( S \) and let \( C' = \{ S - C \mid C \in C \} \).

Then, for every \( K \in \mathcal{P}(S) \) we have \(|C_K| = |C'_K|\).
Proof

We prove the statement by showing the existence of a bijection $f : \mathcal{C}_K \rightarrow \mathcal{C}'_K$. If $U \in \mathcal{C}_K$, then $U = K \cap C$, where $C \in \mathcal{C}$. Then $S - C \in \mathcal{C}'$ and we define $f(U) = K \cap (S - C) = K - C \in \mathcal{C}'_K$. The function $f$ is well-defined because if $K \cap C_1 = K \cap C_2$, then $K - C_1 = K - (K \cap C_1) = K - (K \cap C_2) = K - C_2$.

It is clear that if $f(U) = f(V)$ for $U, V \in \mathcal{C}_K$, $U = K \cap C_1$, and $V = K \cap C_2$, then $K - C_1 = K - C_2$, so $K \cap C_1 = K \cap C_2$ and this means that $U = V$. Thus, $f$ is injective. If $W \in \mathcal{C}'_K$, then $W = K \cap C'$ for some $C' \in \mathcal{C}$. Since $C' = S - C$ for some $C \in \mathcal{C}$, it follows that $W = K - C$, so $W = f(U)$, where $U = K \cap C$. 


Corollary

Let $\mathcal{C}$ be a collection of subsets of a set $S$ and let $\mathcal{C}' = \{ S - C \mid C \in \mathcal{C} \}$. We have $\text{dens}(\mathcal{C}) = \text{dens}(\mathcal{C}')$ and $\text{VCD}(\mathcal{C}) = \text{VCD}(\mathcal{C}')$. 
Theorem

For every collection of sets we have $\text{dens}(\mathcal{C}) \leq VCD(\mathcal{C})$. Furthermore, if $\text{dens}(\mathcal{C})$ is finite, then $\mathcal{C}$ is a VC-class.

Proof: If $\mathcal{C}$ is not a VC-class the inequality $\text{dens}(\mathcal{C}) \leq VCD(\mathcal{C})$ is clearly satisfied. Suppose now that $\mathcal{C}$ is a VC-class and $VCD(\mathcal{C}) = d$. By Sauer-Shelah Theorem we have $\prod_{\mathcal{C}}[m] \leq \phi(d, m)$; then, we obtain $\prod_{\mathcal{C}}[m] \leq \left(\frac{em}{d}\right)^d$, so $\text{dens}(\mathcal{C}) \leq d$.

Suppose now that $\text{dens}(\mathcal{C})$ is finite. Since $\prod_{\mathcal{C}}[m] \leq cm^s \leq 2^m$ for $m$ sufficiently large, it follows that $VCD(\mathcal{C})$ is finite, so $\mathcal{C}$ is a VC-class.
Let $\mathcal{D}$ be a finite collection of subsets of a set $S$. The partition $\pi_{\mathcal{D}}$ was defined as consisting of the nonempty sets of the form 
\[ \{D_1^{a_1} \cap D_2^{a_2} \cap \cdots \cap D_r^{a_r}, \text{where } (a_1, a_2, \ldots, a_r) \in \{0, 1\}^r \} \]

**Definition**

A collection $\mathcal{D} = \{D_1, \ldots, D_r\}$ of subsets of a set $S$ is independent if the partition $\pi_{\mathcal{D}}$ has the maximum numbers of blocks, that is, it consists of $2^r$ blocks.

If $\mathcal{D}$ is independent, then the Boolean subalgebra generated by $\mathcal{D}$ in the Boolean algebra $(\mathcal{P}(S), \{\cap, \cup, -, \emptyset, S\})$ contains $2^{2^r}$ sets, because this subalgebra has $2^r$ atoms. Thus, if $\mathcal{D}$ shatters a subset $T$ with $|T| = p$, then the collection $\mathcal{D}_T$ contains $2^p$ sets, which implies $2^p \leq 2^{2^r}$, or $p \leq 2^r$. 
Definition

Let $\mathcal{C}$ be a collection of subsets of a set $S$. The independence number of $\mathcal{C}$ $I(\mathcal{C})$ is:

$$I(\mathcal{C}) = \sup \{ r \mid \{C_1, \ldots, C_r\} \text{ is independent for some finite } \{C_1, \ldots, C_r\} \subseteq \mathcal{C}\}.$$
Theorem

Let $S, T$ be two sets and let $f : S \rightarrow T$ be a function. If $\mathcal{D}$ is a collection of subsets of $T$ and $\mathcal{C} = f^{-1}(\mathcal{D})$ is the collection $\{ f^{-1}(D) \mid D \in \mathcal{D} \}$, then $I(\mathcal{C}) \leq I(\mathcal{D})$. Moreover, if $f$ is a surjection, then $I(\mathcal{C}) = I(\mathcal{D})$.

Proof: Let $\mathcal{E} = \{ D_1, \ldots, D_p \}$ be an independent finite subcollection of $\mathcal{D}$. The partition $\pi_{\mathcal{E}}$ contains $2^r$ blocks. The number of atoms of the subalgebra generated by $\{ f^{-1}(D_1), \ldots, f^{-1}(D_p) \}$ is not greater than $2^r$. Therefore, $I(\mathcal{C}) \leq I(\mathcal{D})$; from the same supplement it follows that if $f$ is surjective, then $I(\mathcal{C}) = I(\mathcal{D})$. 
Theorem

If \( \mathcal{C} \) is a collection of subsets of a set \( S \) such that \( \text{VCD}(\mathcal{C}) \geq 2^n \), then \( I(\mathcal{C}) \geq n \).

Proof: Suppose that \( \text{VCD}(\mathcal{C}) \geq 2^n \), that is, there exists a subset \( T \) of \( S \) that is shattered by \( \mathcal{C} \) and has at least \( 2^n \) elements. Then, the collection \( \mathcal{H}_t \) contains at least \( 2^{2^n} \) sets, which means that the Boolean subalgebra of \( \mathcal{P}(T) \) generated by \( \mathcal{I}_C \) contains at least \( 2^n \) atoms. This implies that the subalgebra of \( \mathcal{P}(S) \) generated by \( \mathcal{C} \) contains at least this number of atoms, so \( I(\mathcal{C}) \geq n \).