$$m^{1-\alpha} \geq \frac{2(s^d+1)}{-\log(1-\varepsilon)}$$

$$\Longleftrightarrow \quad s^d m^\alpha + m^\alpha \leq -\frac{1}{2}\, m \log(1-\varepsilon).$$

Somit erhalten wir wegen $m^\alpha \geq 1$:

$$\begin{aligned} r &\leq 2^{s^d m^\alpha + 1} \\ &\leq 2^{-\frac{m}{2}\log(1-\varepsilon)} \\ &= (1-\varepsilon)^{-\frac{m}{2}}. \end{aligned}$$

Lemma 5.1 $\Longrightarrow$

$$\begin{aligned} Pr\Big(\sum_{v:\, f_c(v) \neq f_{\bar{c}}^{\,i}(v)} P(v) \geq \varepsilon\Big) &< (1-\varepsilon)^m \cdot r \\ &\leq (1-\varepsilon)^m \cdot (1-\varepsilon)^{-\frac{m}{2}} \\ &= (1-\varepsilon)^{\frac{m}{2}}. \end{aligned}$$

Wegen

$$m \geq \frac{2\cdot \log\frac{1}{\delta}}{-\log(1-\varepsilon)}$$ gilt:

$$\begin{aligned} (1-\varepsilon)^{\frac{m}{2}} &\leq (1-\varepsilon)^{\frac{\log\frac{1}{\delta}}{-\log(1-\varepsilon)}} \\ &= 2^{\frac{\log(1-\varepsilon)\log\frac{1}{\delta}}{-\log(1-\varepsilon)}} \end{aligned}$$

$$= 2^{\log \delta}$$

$$= \delta.$$

Also gilt

$$\Pr\Big(\sum_{v:\, f_c(v) \neq f_{\bar{c}^*}(v)} P(v) \geq \varepsilon \Big) < \delta,$$

so dass unser Satz unter der Voraussetzung, dass das Lemma 5.1 korrekt ist, bewiesen ist.

<u>Bew. von Le 5.1</u>:

Sei

$$B := \Big\{ \bar{c} \in C' \;\Big|\; \sum_{v:\, f_c(v) \neq f_{\bar{c}}(v)} P(v) \geq \varepsilon \Big\}.$$

Betrachten wir $\bar{c} \in B$. Bezeichne $E_{\bar{c}}$ das Ereignis, dass sich $\bar{c}$ bei allen m unabhängig gewählten Beispielen bezüglich der Werte der charakteristischen Funktion genauso wie das zu lernende Konzept c verhält.

Für jedes $\bar{c} \in B$ ist die Wahrscheinlichkeit, dass sich c und $\bar{c}$ für ein zufälliges Beispiel unterscheiden $\geq \varepsilon$.

Da die m Beispiele unabhängig gemäß der

Wahrscheinlichkeitsverteilung P gewählt werden, gilt:

$$Pr(E_{\bar{c}}) < (1-\varepsilon)^m.$$

Hieraus folgt wegen $|B| \leq |C'| = r$

$$Pr\Big(\bigcup_{\bar{c} \in B} E_{\bar{c}}\Big) < (1-\varepsilon)^m \cdot r.$$

∎

5.3 Samplekomplexität von PAC-Lernalgorithmen

Frage:

Gibt es eine Beziehung zwischen der Anzahl der zum Lernen benötigten Beispielen und interessante Eigenschaften der zugrunde liegenden Konzeptklasse?

Ziel: Beantwortung obiger Frage.

Sei $\mathcal{O}\!\!\!\!\top$ ein Lernalgorithmus für eine Konzeptklasse C. Dann heißt die Funktion

$$s : (0,1] \times (0,1] \times \mathbb{N} \to \mathbb{N},$$

wobei $s(\varepsilon, \delta, n)$ die maximale Anzahl der Befragungen des Orakels EXAMPLE durch $\mathcal{O}\!\!\!\!\top$ bei Eingabe ε, δ und n ist, Samplekomplexität des Lernalgorithmus $\mathcal{O}\!\!\!\!\top$. Dabei wird das Maximum über alle zu lernenden Konzepte $c \in C$ und alle Wahrscheinlichkeitsverteilungen P auf $S^{\leq n}$ genommen.

Eine Konzeptklasse C heißt <u>polynomiell samplelernbar</u>, falls ein Lernalgorithmus $\mathcal{A}$ für C und ein Polynom p existieren, so dass $\forall\, (\varepsilon, \delta, n) \in (0,1] \times (0,1] \times \mathbb{N}$ stets

$$s(\varepsilon, \delta, n) \leq p(\tfrac{1}{\varepsilon}, \tfrac{1}{\delta}, n).$$

Seien C eine Konzeptklasse mit zugehörigen Ereignisraum S und $S' \subseteq S$ eine Beispielmenge. Dann besteht $\Pi_C(S')$ genau aus denjenigen Teilmengen von S', die sich als Durchschnitt eines Konzeptes $c \in C$ mit S' schreiben lassen. D.h.,

$$\Pi_C(S') := \{ c \cap S' \mid c \in C \}.$$

Wir sagen C <u>zerlegt</u> $S' \subseteq S$, falls $\Pi_C(S')$ die Potenzmenge von S' ist; d.h.,

$$\Pi_C(S') = 2^{S'}.$$

<u>Beispiel</u>:

Seien

- $S := \mathbb{N}_0$ der Ereignisraum
- $S' := \{0, 1, 2\}$ eine Beispielmenge und

$C := \{ \{0,2,4\}, \{0,4\}, \{2\}, \{4\}, \{0,2\}, \{0,1,2\}, \{1,2\}, \{1,4\}, \{0,1,4\} \}$

Dann zerlegt C die Beispielmenge S'. (überzeugen Sie sich).

◆

Interpretation:

Wenn C die Menge $S' \subseteq S$ zerlegt, dann ist jedes Beispiel $x \in S'$ von den anderen Beispielen in S' unabhängig. D.h., das Wissen $x \in c$ für ein Konzept c sagt nichts über das Enthaltensein irgendeines anderen Beispiels $y \in S'$ in c aus. Je größer die Elementanzahl einer Menge, die von einer Konzeptklasse C zerlegt wird, umso schwieriger ist es, die Konzeptklasse C zu lernen.

Die Vapnik-Chervonenkis (VC) - Dimension $D_{VC}(C)$ von C ist das maximale $d \in \mathbb{N}$, so dass $S' \subseteq S$ mit $|S'| = d$ und C zerlegt S' existiert.

Bemerkung:

- Um zu beweisen, dass die VC-Dimension einer Konzeptklasse C mindestens d ist, genügt es eine Beispielmenge S' der Größe

d zu konstruieren und zu zeigen, dass C die Beispielmenge S' zerlegt.

- Um zu zeigen, dass die VC-Dimension einer Konzeptklasse C höchstens d ist muss gezeigt werden, dass jede Beispielmenge der Größe $d+1$ nicht von C zerlegt wird.

- Um für eine Beispielmenge S' zu zeigen, dass diese nicht von einer Konzeptklasse C zerlegt wird, genügt es, für eine Teilmenge $S'' \subseteq S'$ zu beweisen, dass kein Konzept $c \in C$ mit $c \cap S' = S''$ existiert.

<u>Beispiel:</u>
Wir betrachten zwei geometrische Konzeptklassen und berechnen ihre VC-Dimension.

a) Seien
- die reele Gerade der Ereignisraum S_1 und
- die Menge der abgeschlossenen Intervalle auf S_1 die Konzeptklasse C_1.

$\Longrightarrow$

C_1 zerlegt jede Beispielmenge bestehend aus zwei beliebigen Punkten auf S_1

aber

$\nexists$ Beispielmenge S'_1 bestehend aus drei Punkten auf S_1, die C_1 zerlegt.

$\Longrightarrow$

VC-Dimension der Konzeptklasse C_1 ist zwei.

b) Seien

- die reele Ebene der Ereignisraum S_2 und
- die Menge der Halbebenen von S_2 die Konzeptklasse C_2.

Sei S_2' eine Menge von drei beliebigen Punkten in S_2, die nicht auf einer Geraden liegen.

$\Longrightarrow$

C_2 zerlegt die Beispielmenge S_2'.

$\Longrightarrow$

VC-Dimension der Konzeptklasse C_2 ist $\geq$ drei

<u>Beh.</u>: VC-Dimension von C_2 ist < vier.

<u>Bew.</u>:

Betrachte vier beliebige Punkte S_2' in der Ebene.
Drei Fälle können eintreten

1) Mindestens drei der Punkte liegen auf derselben Geraden.

2) Es liegen maximal zwei Punkte auf derselben Gerade und alle vier Punkte liegen auf der von ihnen definierten konvexen Hülle.

3) Drei der vier Punkte definieren die konvexe Hülle und der vierte Punkt liegt im Inneren der konvexen Hülle.

Wir diskutieren die drei Fälle nacheinander.

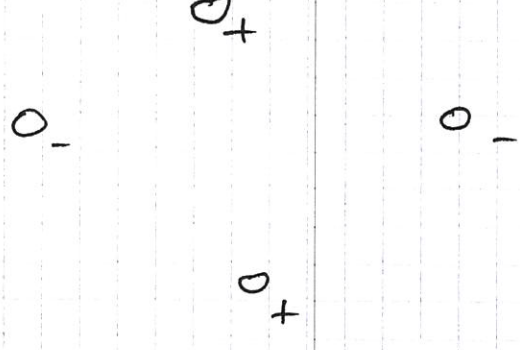
<u>1. Fall:</u>

∄ Halbebene, die von drei Punkten auf derselben Gerade die beiden äußersten Punkte und nicht den inneren Punkt enthält.

$\Rightarrow$

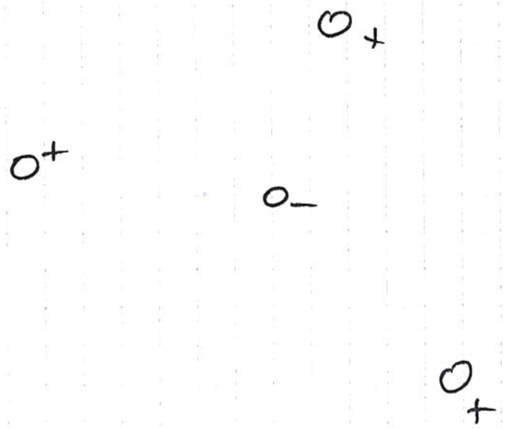
C_2 zerlegt Beispielmenge S_2' nicht.

<u>2. Fall:</u>



Betrachte Beispielmenge S_2'', die aus zwei auf der konvexen Hülle gegenüberliegenden Punkten besteht. ∄ Halbebene, ~~die~~ die die Punkte in S_2'' und nicht mindestens einen Punkt in $S_2' \setminus S_2''$ enthält. $\Rightarrow$ C_2 zerlegt die Beispielmenge S_2' nicht.

3. Fall



Betrachte die Beispielmenge S_2^4, die exakt die drei Punkte auf der konvexen Hülle enthält.
$\nexists$ Halbebene, die S_2^4 und nicht den Punkt in $S_2^1 \setminus S_2^4$ enthält.

$\Rightarrow$ C_2 zerlegt Beispielmenge S_2^1 nicht ■

Insgesamt gilt

$$D_{VC}(C_2) = 3.$$

◆

Übung:
Seien die euklidische Ebene $\mathbb{R}^2$ der Ereignisraum S und die Menge der zu den Koordinatenachsen parallelen Rechtecke die Konzeptklasse C. Zeigen Sie, dass vier die VC-Dimension von C ist.

Lemma 5.2

Seien S eine endliche Menge, C eine Konzeptklasse mit zugehörigen Ereignisraum S und

$d := D_{VC}(C)$. Dann gilt

$$2^d \leq |C| \leq (|S|+1)^d.$$

Beweis:

Da C eine Menge $S' \subseteq S$ mit $|S'| = d$ zerlegt und die Anzahl der verschiedenen Teilmengen von S' gerade 2^d ist, muss C mindestens 2^d verschiedene Konzepte enthalten.

Die obere Schranke beweisen wir mittels Induktion über $|S|$.

$|S| = 1$:

Dann existieren nur die Konzeptklassen

$$\{\emptyset\}, \{S\} \text{ und } \{\emptyset, S\}.$$

Die Konzeptklassen $\{\emptyset\}$ und $\{S\}$ haben die VC-Dimension 0. $\{\emptyset, S\}$ hat VC-Dimension 1. Also gilt offensichtlich die obere Schranke.

Annahme:

$|C| \leq (|S|+1)^d$ für alle Ereignisräume S mit $|S| = k$, $k \geq 1$.

$k \leadsto k+1$:

Sei $|S| = k+1$. Betrachten wir $x \in S$ beliebig aber fest. Seien

$S_1 := S \setminus \{x\}$,

$C_1 := \{ c_1 \in C \mid \exists\, c \in C \setminus \{c_1\} \text{ mit } c_1 = c \cup \{x\} \}$

$C_2 := C \setminus C_1$.

Idee:

Herleitung von oberen Schranken für $|C_1|$ und $|C_2|$. Deren Summe ist dann eine obere Schranke für $|C|$.

Konstruktion $\Rightarrow$

Zwei verschiedene Konzepte in C_2 können sich nicht nur bezüglich x unterscheiden.

$\Rightarrow$

Die Konzepte in C_2 können mittels Beispiele in S_1 unterschieden werden.

$\Rightarrow$

Wir können C_2 als Konzeptklasse mit zugehöriger Ereignismenge S_1 interpretieren, ohne dass sich hierdurch die Elementanzahl von C_2 ändern würde.

Wegen $C_2 \subseteq C$ kann C_2 keine größere Menge zerlegen als C.

$\Rightarrow$

$D_{VC}(C_2) \leq d$.

Wegen $|S_1| = k$ ist die Induktionsannahme anwendbar.

$\Longrightarrow$

$$|C_2| \leq (|S_1|+1)^d.$$

Da alle Konzepte in C_1 x enthalten, unterscheiden sich verschiedene Konzepte in C_1 durch Beispiele aus S_1.

<u>Ziel</u>: Beweis, dass $D_{VC}(C_1) < d$.

<u>Annahme</u>: C_1 zerlegt $S' \subseteq S_1$ mit $|S'| \geq d$.

Konstruktion von C_1 und C_2 $\Longrightarrow$

C zerlegt $S' \cup \{x\}$.

Wegen

$$|S' \cup \{x\}| \geq d+1 \quad \text{und} \quad D_{VC}(C) = d$$

kann dies nicht sein.

$\Longrightarrow$

Teilmengen von S_1, die C_1 zerlegt, enthalten höchstens $d-1$ Elemente.

Induktionsannahme $\Longrightarrow$

$$|C_1| \leq (|S_1|+1)^{d-1}.$$

Insgesamt erhalten wir

$$|C| = |C \setminus C_1| + |C_1|$$

$$= |C_2| + |C_1|$$

$$
\begin{aligned}
&\leq (|S_1|+1)^d + (|S_1|+1)^{d-1} \\
&= (k+1)^d + (k+1)^{d-1} \\
&= (k+2)(k+1)^{d-1} \\
&\leq (k+2)^d \\
&= (|S|+1)^d !
\end{aligned}
$$

■

Ziel:

Herausarbeitung eines Zusammenhanges zwischen VC-Dimension und polynomieller Samplelernbarkeit.

Nachfolgend ist stets $\Sigma = \{0,1\}$.

Seien C eine Konzeptklasse mit zugehörigen Ereignisraum Σ^* und $c \in C$. Die Projektion $c^{\leq n}$ von c auf $\Sigma^{\leq n}$ ist definiert durch

$$c^{\leq n} := c \cap \Sigma^{\leq n}.$$

Die Projektion $C^{\leq n}$ von C auf $\Sigma^{\leq n}$ ist dann definiert durch

$$C^{\leq n} := \{ c^{\leq n} \mid c \in C \}.$$

Die Funktion $d : \mathbb{N} \to \mathbb{N}$ mit

$$d(n) = D_{VC}(C^{\leq n}) \quad \forall\, n \in \mathbb{N}$$

heißt asymptotische VC-Dimension der Konzeptklasse C. Diese bezeichnen wir mit $D^a_{VC}(C)$.

Wir sagen, C hat polynomielle VC-Dimension, falls ein Polynom p existiert, so dass

$$D^{a}_{VC}(C)(n) = O(p(n)) \quad \forall n \in \mathbb{N}.$$

Folgendes Lemma folgt direkt aus Lemma 5.2:

Lemma 5.3

Seien $n \in \mathbb{N}$ und $d = D_{VC}(C^{\leq n})$. Dann gilt

$$2^{d} \leq |C^{\leq n}| \leq 2^{(n+1)d}.$$

Beweis:

Nach Substitution von $C^{\leq n}$ für C und von $\Sigma^{\leq n}$ für S in Lemma 5.2 erhalten wir wegen $|\Sigma^{\leq n}| \leq 2^{n+1} - 1$:

$$2^{d} \leq |C^{\leq n}| \leq (2^{n+1} - 1 + 1)^{d} = 2^{(n+1)d}.$$

■

Ziel:

Entwicklung eines Lernalgorithmus für Konzeptklassen mit Ereignisraum Σ^*, dessen Samplekomplexität mit Hilfe der asymptotischen VC-Dimension der Konzeptklasse ausgedrückt werden kann.

Satz 5.4

Sei C eine Konzeptklasse mit zugehörigen Ereignisraum Σ^*. Dann existiert ein Lernalgorithmus $\mathcal{A}_{VC}$ für C mit Sample Komplexität

$$s(\varepsilon, \delta, n) = \left\lceil \frac{1}{\varepsilon}\left((n+1) D_{VC}(C^{\leq n}) \ln 2 + \ln \frac{1}{\delta} \right) \right\rceil .$$

Beweis:

Betrachten wir folgenden Lernalgorithmus für C:

Algorithmus $\mathcal{A}_{VC}$

Eingabe: $\varepsilon, \delta \in (0,1]$, $n \in \mathbb{N}$.

Ausgabe: $g \in C$.

Methode:

$m := \frac{1}{\varepsilon}\left((n+1) D_{VC}(C^{\leq n}) \ln 2 + \ln \frac{1}{\delta} \right)$;
Befrage EXAMPLE m-mal;
(* Sei Z die Menge der gesehenen Beispiele. *)
Ermittle ein mit Z konsistentes Konzept $g \in C$;
Ausgabe := g.

Ziel:

Beweis, dass der Algorithmus $\mathcal{A}_{VC}$ die erforderlichen Eigenschaften besitzt.

Sei $c \in C$ das zu lernende Konzept. Dann ist

$$Pr\left(\sum_{v: f_c(v) \neq f_g(v)} P(v) < \varepsilon \right) \geq 1 - \delta$$

zu beweisen. Der Beweis verläuft ähnlich zum Beweis des Lemmas 5.1. Sei

$$B^{\leq n} := \left\{ h^{\leq n} \in C^{\leq n} \;\middle|\; \sum_{v \in \Sigma^{\leq n}: f_c(v) \neq f_{h^{\leq n}}(v)} P(v) \geq \varepsilon \right\}.$$

Für jedes $g^{\leq n} \in B^{\leq n}$ ist die Wahrscheinlichkeit, dass sich $f_c(v)$ und $f_{g^{\leq n}}(v)$ für ein zufällig gewähltes Beispiel $v \in \Sigma^{\leq n}$ unterscheiden, mindestens ε.

Bezeichne $E_{g^{\leq n}}$ das Ereignis, dass $g^{\leq n}$ bei allen m unabhängig gewählten Beispielen sich bzgl. der Werte der charakteristischen Funktion genauso wie c verhält. Dann gilt

$$Pr(E_{g^{\leq n}}) \leq (1-\varepsilon)^m.$$

Da $c \notin B^{\leq n}$ und somit $|B^{\leq n}| < |C^{\leq n}|$ erhalten wir

$$Pr\left(\bigcup_{g^{\leq n} \in B^{\leq n}} E_{g^{\leq n}} \right) < |C^{\leq n}| \cdot (1-\varepsilon)^m.$$

Idee:
Wähle m groß genug, so dass $|C^{\leq n}| \cdot (1-\varepsilon)^m \leq \delta$.

Lemma 5.3 $\Rightarrow$

$$|C^{\leq n}| \leq 2^{(n+1) D_{VC}(C^{\leq n})}.$$

Somit genügt es m derart zu wählen, dass

$$2^{(n+1) D_{VC}(C^{\leq n})} (1-\varepsilon)^m \leq \delta.$$

$$\Leftrightarrow (n+1) D_{VC}(C^{\leq n}) \ln 2 + m \cdot \ln(1-\varepsilon) \leq \ln \delta.$$

Wegen $\ln(1+x) \leq x$ genügt es zu zeigen, dass

$$(n+1) D_{VC}(C^{\leq n}) \ln 2 - m\varepsilon \leq \ln \delta$$

$$\Leftrightarrow \quad m \geq \frac{1}{\varepsilon}\left((n+1) D_{VC}(C^{\leq n}) \ln 2 + \ln \frac{1}{\delta}\right).$$

Also gilt für $m \geq \frac{1}{\varepsilon}\left((n+1) D_{VC}(C^{\leq n}) \ln 2 + \ln \frac{1}{\delta}\right)$

$$\Pr\Big(\sum_{v \in \Sigma^{\leq n}: f_c(v) \neq f_{g^{\leq n}}(v)} P(v) < \varepsilon\Big) \geq 1-\delta,$$

wobei g konsistent mit allen m unabhängig gewählten Beispielen ist.

■

Nach Einsetzen von $p(n)$ für $D_{VC}(C^{\leq n})$ in $\frac{1}{\varepsilon}\left((n+1) D_{VC}(C^{\leq n}) \ln 2 + \ln \frac{1}{\delta}\right)$ für ein Polynom p erhalten wir wiederum ein Polynom. Also impliziert Satz 5.4 direkt folgendes Korollar:

<u>Korollar 5.1</u>

Sei C eine Konzeptklasse mit zugehörigem Ereignisraum Σ^* und polynomieller VC-Dimension. Dann ist C polynomiell samplelernbar.

Folgende Frage drängt sich auf:

Wie gut ist Satz 5.4?

Da wir in obiger Analyse als obere Schranke für $|B^{\leq n}|$ einfach $|C^{\leq n}|-1$ genommen haben, könnte man sich zunächst fragen, ob wir obige Analyse des Lernalgorithmus verbessern können. In der Tat kann man durch eine kompliziertere Analyse in obiger oberen Schranke n durch eine Konstante ersetzen.

siehe: M.J. Kearns, U.V. Vazirani, An Introduction ...

Des Weiteren ist es zur Beantwortung obiger Frage sinnvoll, sich Gedanken über untere Schranken für die Samplekomplexität einer Konzeptklasse C in Abhängigkeit seiner asymptotischen VC-Dimension zu machen.

<u>Satz 5.5</u>

Jeder PAC-Lernalgorithmus für eine Konzeptklasse C hat bei einem Fehlerparameter $\varepsilon < \frac{1}{4}$ Samplekomplexität $s(\varepsilon, \delta, n) \geq \frac{1}{8\varepsilon} D_{VC}(C^{\leq n})$.

<u>Beweis:</u>

<u>Idee</u>:

Konstruiere auf dem Ereignisraum S eine Wahrscheinlichkeitsverteilung P, die jeden

PAC-Lernalgorithmus zwingt, viele Beispiele zu betrachten.

Durchführung:

Seien $d := D_{VC}(C^{\leq n})$ und $S' := \{x_1, x_2, \dots, x_d\}$ eine Beispielmenge, die von $C^{\leq n}$ zerlegt wird.

Betrachten wir die Wahrscheinlichkeitsverteilung P, die jedem Beispiel in S' die Wahrscheinlichkeit $\frac{1}{d}$ und jedem anderen Beispiel die Wahrscheinlichkeit 0 zuordnet.

Sei c das zu lernende Konzept.

Annahme:

Ein Lernalgorithmus $\mathcal{O}\!\mathcal{T}$ befragt bei obiger Wahrscheinlichkeitsverteilung m-mal EXAMPLE.

Für den Beweis der unteren Schranke können wir O.B.d.A. annehmen, dass EXAMPLE stets ein neues Beispiel aus S' auswählt.

Sei $S'' \subseteq S'$ mit $|S''| = m$ die Menge der Beispiele, die $\mathcal{O}\!\mathcal{T}$ aufgrund seiner Befragungen von EXAMPLE kennt.

$C^{\leq n}$ zerlegt die Beispielmenge S' $\Longrightarrow$

$\mathcal{O}\!\mathcal{T}$ hat bezüglich jedem Beispiel $x \in S' \setminus S''$ keine Information, ob $x \in c^{\leq n}$ oder nicht.

$\Longrightarrow$

Für jede der möglichen 2^{d-m} Teilmengen von $S' \setminus S''$ gibt es in $C^{\leq n}$ ein mit S' konsistentes Konzept, das von den Beispielen in $S' \setminus S''$ exakt diese Teilmenge enthält.

Sei $B \subseteq C$ eine Menge von Konzepten, die für jede Teilmenge von $S' \setminus S''$ exakt ein solches Konzept enthält.

$\Longrightarrow$

$|B| = 2^{d-m}.$

Sei h die Ausgabe von $\mathcal{O}$.

Für jedes Beispiel $x \in S' \setminus S''$ gilt für exakt der Hälfte der Konzepte in B, dass diese und h sich unterschiedlich bzgl. x verhalten.

$\Longrightarrow$

$$\sum_{c \in B} \sum_{x \in S' \setminus S'' :\, f_c(x) \neq f_h(x)} P(x) = \frac{(d-m)\, 2^{d-m}}{2d}$$

$\Longrightarrow$

$\exists\, c_0 \in B$, so dass

$$\sum_{x \in S' \setminus S'' :\, f_{c_0}(x) \neq f_h(x)} P(x) \geq \frac{d-m}{2d}.$$

Für $m \leq \frac{d}{2}$ gilt dann

$$\sum_{x \in S' \setminus S'' : f_{c_0}(x) \neq f_h(x)} P(x) \geq \frac{1}{4}.$$

Da sich EXAMPLE bezüglich S'' für alle zu lernenden Konzepten in B gleich verhält, können wir aus B jedes Konzept als das zu lernende Konzept auswählen, ohne dass sich hierdurch die Ausgabe von $\mathcal{O}\mathcal{I}$ ändern würde. Also gilt

$$s(\varepsilon, \delta, n) > \frac{D_{VC}(C^{\leq n})}{2}.$$

<u>Idee</u>:

Um in die untere Schranke die gewünschte Abhängigkeit von dem Fehlerparameter ε hineinzubekommen, modifizieren wir obige Wahrscheinlichkeitsverteilung P, so dass die Wahrscheinlichkeiten von ε abhängen.

<u>Durchführung</u>:

Sei die Wahrscheinlichkeitsverteilung Q definiert durch

1) $Q(x_1) := 1 - 4\varepsilon$,
2) $Q(x_i) := \dfrac{4\varepsilon}{d-1}$ $\quad 2 \leq i \leq d$ und
3) $Q(x) := 0 \quad \forall x \in S \setminus S'$.

Bei der Wahrscheinlichkeitsverteilung Q be=

trägt die erwartete Anzahl von Befragungen an EXAMPLE, damit von EXAMPLE ein Beispiel aus $\{x_2, x_3, \ldots, x_d\}$ gewählt wird $\frac{1}{4\varepsilon}$.

$\Rightarrow$

Mindestens $\frac{m}{4\varepsilon}$ Befragungen von EXAMPLE werden benötigt, damit der Algorithmus $\mathcal{A}$ mindestens m unterschiedliche Beispiele kennenlernt.
Für $m := \frac{d}{2}$ ergibt dies die Anzahl $\frac{d}{8\varepsilon}$.

Analog zu oben zeigt man nun

$$S(\varepsilon, \delta, n) > \frac{D_{VC}(C^{\leq n})}{8\varepsilon}$$ Übung

■