# CSCI B609: “Foundations of Data Science”

## Lecture 11/12: VC-Dimension and VC-Theorem

Slides at http://grigory.us/data-science-class.html


**Grigory Yaroslavtsev**

**http://grigory.us**

# Intro to ML

- Classification problem
  - Instance space $:\{0,1\}$ or $\mathbb{R}$ (feature vectors)
  - Classification: come up with a mapping $\to\{0,1\}$
- Formalization:
  - Assume there is a probability distribution over
  - $*$ = “target concept” (set $* \subseteq$ of positive instances)
  - Given labeled i.i.d. samples from produce $\subseteq$
  - **Goal:** have agree with $*$ over distribution
  - Minimize: $(\ ) = \Pr[\ \Delta\ *\ ]$

# Intro to ML

- Training error
  - $= $ labeled sampled (pairs $( , )$, $\in$ , $\in \{0,1\}$)
  - Training error: $( ) = \frac{| \cap ( \Delta ^*)|}{| |}$
- "Overfitting": low training error, high true error
- Hypothesis classes:
  - H: collection of subsets of called hypotheses
    - If $= \mathbb{R}$ could be all intervals $\{[ , ], \le \}$
    - If $= \mathbb{R}$ could be linear separators: $\{\{ \in \mathbb{R} \mid \cdot \ge {}_0\} / \in \mathbb{R} , {}_0 \in \mathbb{R}\}$
- If is large enough (compared to some property of H) then overfitting doesn't occur

# Overfitting and Uniform Convergence

- **PAC learning (agnostic)**: For $\;,\;\; > 0$ if

$$| \;\; | \geq 1/2\;\;^{2}(\ln|\;\;| + \ln 2/\;\;)$$

then with probability $1 - \;\;$:

$$\forall \;\; \in \boldsymbol{H}: |\;\;(\;\;) - \;\;(\;\;)| \leq$$

- Size of the class of hypotheses can be very large
- Can also be infinite, how to give a bound then?
- We will see ways around this today

# VC-dimension

- VC-dim( ) $\leq \ln | \quad |$
- Consider database age vs. salary
- Query: fraction of the overall population with ages 35 – 45 and salary \$$(50 - 70)$K
- How big a database can answer with $\pm$ error
- 100 ages $\times$ 1000 salaries $\Rightarrow 10^{10}$ rectangles
- $1/2\ {}^{2}(10 \ln 10 + \ln 2/\quad)$ samples suffice
- What if we don't want to discretize?

# VC-dimension

- **Def.** Concept class **shatters** a set if $\forall \subseteq$ there is $\mathrm{h} \in$ labeling positive and $\mathrm{A} \setminus$ negative
- **Def. VC-dim**( ) = size of the largest shattered set
- Example: axis-parallel rectangles on the plane
  - 4-point diamond is shattered
  - No 5-point set can be shattered
  - VC-dim(axis-parallel rectangles) = 4
- **Def.** $[\ ] = \{h \cap \ : h \in \ \}$ **=** set of labelings of the points in by functions in
- **Def. Growth function** $(\ ) = \max_{|\ |=} /\ [\ ]/$

# Growth function & uniform convergence

- **PAC learning via growth function**: For $,\quad > 0$ if

$$|\quad| = \quad \geq 8/\quad{}^{2}(\ln|2\quad(2\quad)| + \ln 1/\quad)$$

then with probability $1 -$ :

$$\forall \quad \in \boldsymbol{H}: |\quad(\quad) - \quad(\quad)| \leq$$

- **Thm (Sauer's lemma).** If VC-dim($\boldsymbol{H}$) = then:

$$(\quad) \leq \sum_{=0}(\quad) \leq \left(\frac{}{}\right)$$

- For half-planes, $\mathrm{VC} - \dim = 3,\quad (\quad) = (\quad{}^{2})$

# Sauer's Lemma Proof

- Let $=$ -dim( ) we'll show that if $| \quad | = \quad :$

$$| \quad [ \quad ]| \leq \binom{}{\leq} = \sum_{=0} \binom{}{}$$

- $\binom{}{\leq} = \binom{-1}{\leq} + \binom{-1}{\leq \quad -1}$

Proof (induction by set size):

- $\setminus \{ \quad \}$: by induction $| \quad [ \quad \setminus \{ \quad \}]| \leq \binom{-1}{\leq}$
- $| \quad [ \quad ]| - | \quad [ \quad \setminus \{ \quad \}]| \leq \binom{-1}{} ?$

$$|\;[\;]| - |\;[\;\setminus\{\;\}]| \le \binom{\;-1}{\le\;-1}$$

- If $[\;] > [\;\setminus\{\;\}]$ then it is because of the sets that differ only on so let's pair them up
- For $h \in [\;]$ containing let $(\;) = h \setminus \{\;\}$
  $= \{h \in [\;]:\; \in h \qquad (\;) \in [\;]\}$
- Note: $|\;[\;]| - |\;[\;\setminus\{\;\}]| = |\;|$
- What is the VC-dimension of ?
  - If VC-dim$(\;) = \;'$ then $\subseteq \;\setminus\{\;\}$ of $\;'$ is shattered
  - All $2^{\;'}$ subsets of are 0/1 extendable on
  - $\ge \;' + 1 \Rightarrow$ VC-dim$(\;) \le \; - 1 \Rightarrow$ apply induction

# Examples

- Intervals of the reals:
  - Shatter 2 points, don’t shatter 3 $\Rightarrow$ -dim $= 2$
- Pairs of intervals of the reals:
  - Shatter 4 points, don’t shatter 5 $\Rightarrow$ -dim $= 4$
- Convex polygons
  - Shatter any points on a circle $\Rightarrow$ -dim $= \infty$
- Linear separators in dimensions:
  - Shatter $+ 1$ points (unit vectors + origin)
  - Take subset S and set $= 0$ if $\in$ :

separator $\leq 0$

# VC-dimension of linear separators

No set of $+2$ points can be shattered

- **Thm (Radon).** Any set $\subseteq \mathbb{R}$ with $|\quad| = \quad + 2$

can be partitioned into two subsets $,$ s.t.:

$$\text{Convex}(\quad) \cap \text{Convex}(\quad) \neq \emptyset$$

- Form $\times (\quad + 2)$ matrix A, columns = points in
- Add extra all-1 row ⇒ matrix B
- $= ({}_{1}, {}_{2}, \ldots, {}_{+2})$, non-zero vector: $= 0$
- Reordering: ${}_{1}, {}_{2}, \ldots, \geq 0, {}_{+1}, \ldots, {}_{+2} < 0$

$\sum$

# Radon's Theorem (cont.)

- , $=$ i-th columns of and
- $\sum_{=1} | \quad | = \sum_{=\ +1}^{+2} | \quad |$

  $-\sum_{=1} | \quad | = \sum_{=\ +1}^{+2} | \quad |$

  $-\sum_{=1} | \quad | = \sum_{=\ +1}^{+2} | \quad | = 1$
- Convex combinations of two subsets intersect

# Growth function & uniform convergence

- **PAC learning via growth function**: For $, > 0$ if

$$| \; | = \; \geq 8/ \; ^2 ( \ln |2 \; (2 \; )| + \ln 1/ \; )$$

then with probability $1 - \;$:

$$\forall \; \in H: | \; ( \; ) - \; ( \; )| \leq$$

- Assume **event A**:

$$\exists \; \in H: | \; ( \; ) - \; ( \; )| >$$

- Draw $'$ of size $,$ **event B**:

$$\exists \; \in H: \; | \; ( \; ) - \; ( \; )| >$$

$$| \; '( \; ) - \; ( \; )| < \; /2$$

$$[\quad] \geq \Pr[\boxed{f()}]/2$$

- **Lem.** If $= \Omega(1/\ ^2)$ then $[\quad] \geq \Pr[\boxed{f()}]/2$.
- **Proof:**

$$[\quad] \geq \Pr[\ ,\ ] = \Pr[\quad]\Pr[\boxed{f()}/\ ]$$

- Suppose occurs:

$$\exists \quad \in \boldsymbol{H}: |\quad (\ ) - \quad (\ )| >$$

- When we draw $'$:

$$_{'}[\quad _{'}(\ )] = \quad (\ )$$

- By Chernoff:

$$_{'}[|\quad _{'}(\ ) - \quad (\ )| < \ /2] \geq \frac{1}{2}$$

# VC-theorem Proof

- Suffices to show that $[\ ] \leq \ /2$
- Consider drawing $2$ samples $''$ and then randomly partitioning into $'$ and
- $*$ *:* same as for such $(\ ',\ ) \Rightarrow \Pr[\ *]$ $= \Pr[\ ]$
- **Will show:** $\forall$ fixed $''$ $_{,S'}[\ */\ '']$ is small
- **Key observation:** once $''$ is fixed there are only $/\ [\ '']/ \leq \ (2\ )$ events to care about
- Suffices: for every fixed $h \in \ [\ '']$:

# VC-theorem Proof (cont.)

- Randomly pair points in $''$ into $( \quad , \quad )$ pairs
- With prob. ½: $\rightarrow \; , \quad \rightarrow \; '$ or $\rightarrow \; ', \quad \rightarrow$
- Diff. between $( \; )$ and $'( \; )$ for $= 1, \ldots,$
- Only changes if mistake on only one of $( \quad , \quad )$
  - With prob. ½ difference changes by $\pm 1$
  - By Chernoff:

$$Pr\left[\left| \quad ( \; ) - \quad '( \; ) \right| > \frac{\quad}{4}\right] = \quad ^{-\Omega( \; ^2 \; )}$$

- $\quad ^{-\Omega( \; ^2 \; )} \leq$ for from the Thm. statement