From VC-Dimension to Learning Results

- Let $X$ be a domain of examples, $D$ a distribution over $X$, $C$ a concept class over $X$, and $f \in C$ the target concept.
- The set of error regions relative to $f$: $\Delta(f) = \{f \Delta c \mid c \in C\}$ where $f \Delta c$ stands for the symmetric difference of the concepts.
- Heavy-weighted error regions: $\Delta_{\epsilon}(f) = \{r \in \Delta(f) \mid Pr_D[r] > \epsilon\}$
- Useful samples:
  We say that a sample $S = \{x_1, \ldots, x_m\}$ of $m$ examples is an $\epsilon$-net if for all $r \in \Delta_{\epsilon}(f)$ we have $S \cap r \neq \emptyset$.
- Fact: if $S$ is an $\epsilon$-net and $h \in C$ is consistent with $S$, then
  $$\text{error}(h) = Pr_{x \in D}[h(x) \neq f(x)] \leq \epsilon.$$

- **Theorem:** Let $VCD(C) = d$.
  If an IID sample $S$ is of size
  $$m \geq \max\left\{\frac{6}{\epsilon} \ln \frac{2 \cdot 8d}{\delta}, \frac{8d}{\epsilon} \ln \frac{30}{\epsilon}\right\},$$
  then, with probability at least $1 - \delta$, $S$ is an $\epsilon$-net.
- **Corollary:** Assume an algorithm $A$ is guaranteed to find a consistent hypothesis in $C$.
  Run $A$ with an IID sample size $m$ as in theorem and let its output be hypothesis $h$.
  Then with probability at least $1 - \delta$, $\text{error}(h) \leq \epsilon$.
- **Alternative Way to State Result:**
  with probability at least $1 - \delta$,
  $$\text{error}(h) \leq \epsilon = \frac{2}{m} \left(\log \frac{2}{\delta} + d \log \frac{2em}{d}\right).$$

- **Probability Experiment 1:** Draw IID sample $S_1$ of $m$ examples.
  - **Event A:** $S_1$ is not an $\epsilon$-net.

- **Probability Experiment 2:** Draw two IID samples $S_1, S_2$ each of $m$ examples.
  - **Event B:** event A happens and for some $r \in \Delta_{\epsilon}(f)$ such that $r \cap S_1 = \emptyset$ we have $|r \cap S_2| \geq \frac{em}{2}$.
  - **Claim 1:** If $m \geq \frac{8}{\epsilon} \ln 2$, then $Pr[A] \leq 2Pr[B]$.

- **Probability Experiment 3:**
  - Draw IID sample $S$ of $2m$ examples.
  - Randomly permute the order of examples.
  - Let $S_1$ be first $m$ examples and $S_2$ the rest.
  - **Fact:** same distribution as in experiment 2.

- **Event B:** event A happens and for some $r \in \Delta_{\epsilon}(f)$ such that $r \cap S_1 = \emptyset$ we have $|r \cap S_2| \geq \frac{em}{2}$.
  - **Same as:** $\exists r \in \Delta_{\epsilon}(f)$ such that $|r \cap S_2| \geq \frac{em}{2}$ and $r \cap S_1 = \emptyset$.

- **Probability Experiment 4:**
  (parameterized by sample $S$ of $2m$ examples)
  - Randomly permute the order of examples.
  - Let $S_1$ be first $m$ examples and $S_2$ the rest.
  - **Event B’(S):** (same as B but for fixed $S$) $\exists \tilde{r} \in \Pi_{\Delta_{\epsilon}(f)}(S)$ such that $|\tilde{r} \cap S_2| \geq \frac{em}{2}$ and $\tilde{r} \cap S_1 = \emptyset$. 

• **Claim 2:** \( \Pr[B] \leq \max_S \Pr[B'(S)] \).

• **Claim 3:**
  \( \Pr[B'(S)] \leq |\Pi_{\Delta(f)}(S)| \cdot \max_{\hat{r}} \Pr[\hat{r} \subseteq S_2] \)
  where \( \hat{r} \in \Pi_{\Delta(f)}(S) \) and \( |\hat{r}| \geq \frac{em}{2} \).

• **Claim 4:** For any \( \hat{r} \in \Pi_{\Delta(f)}(S) \) and such that \( |\hat{r}| \geq \frac{em}{2} \):
  \( \Pr[\hat{r} \subseteq S_2] \leq \left( \frac{1}{2} \right)^{\frac{em}{2}} \).

• **Claim 5:** \( |\Pi_{\Delta(f)}(S)| \leq \Phi_d(2m) \leq \left( \frac{2em}{d} \right)^d \).

• **Proof of Theorem:**
  \[
  \Pr[A] \leq 2\Pr[B] \leq 2\Pr[B'(S)] \\
  \leq 2 \left( \frac{2em}{d} \right)^d \left( \frac{1}{2} \right)^{\frac{em}{2}} \leq 2\delta = \delta \\
  \text{last step uses: } m \geq \max \left\{ \frac{6}{\epsilon} \ln \frac{2}{\delta}, \frac{8d}{\epsilon} \ln \frac{30}{\epsilon} \right\}
  \]