Occam Algorithms

• Let $X$ be a domain of examples and $C, H$ concept classes over $X$.

• **Theorem 0:** Let $A$ be a learning algorithm such that $\forall c \in C$ $A$ takes a sample of $m$ examples and outputs a hypothesis $h \in H$ consistent with $S$.
Then when using a sample of size
$$m \geq \frac{1}{\epsilon} \ln \frac{|H|}{\delta},$$
A is a PAC learning algorithm for $C$ using $H$.

• Let $X = \cup_{n\geq1}X_n$ be a stratified domain and $C = \cup_{n\geq1}C_n$, $H = \cup_{n\geq1}H_n$, be stratified concept classes over $X$.

• Let $A$ be a learning algorithm that takes a sample $S$ and returns a hypothesis $h$. Let $H_{n,m} = \{h|h$ is output by $A$ when learning $c \in C_n$ with a sample of $m$ examples}.

• **Theorem 1:** If $\forall n, \forall c \in C_n, \forall S$ s.t. $|S| = m$, $A$ outputs a hypothesis $h \in H_{n,m}$ consistent with $S$.
Then when using a sample of size
$$m \geq \frac{1}{\epsilon} \ln \frac{|H_{n,m}|}{\delta},$$
A is a PAC learning algorithm for $C$ using $H$.

• Note that, in the above, a constraint between $m$ and $|H_{n,m}|$ must be satisfied.

Compression Algorithms

• **Definition:** For $\alpha \geq 0$ and $0 \leq \beta < 1$.
Algorithm $A$ is an $(\alpha, \beta)$-Occam algorithm for learning $C$ by $H$ if on any sample $S \subseteq X_n$ of $m$ examples labeled by $c \in C_n$ $A$ outputs $h \in H_n$ consistent with $S$ and such that
$$size(h) \leq [n \cdot size(c)]^{\alpha}m^{\beta}.$$  
• In the above, $size(x)$ measures the number of bits used to represent $x$ in the respective representation language.

• **Theorem 2:** If $A$ is an $(\alpha, \beta)$-Occam algorithm for learning $C$ by $H$.
Then when using a sample of size
$$m \geq \max \left\{ \frac{2}{\epsilon} \ln \frac{1}{\delta}, \left(\frac{2[n \cdot size(c)]^{\alpha}}{\epsilon}\right)^{\frac{1}{1-\beta}} \right\},$$
A is a PAC learning algorithm for $C$ using $H$.

• Fix any learning algorithm $L$ that takes an ordered sample $B$ as input and produces a hypothesis $hyp(B)$ based on the sample.

• Algorithm $A$ is a compression algorithm for concept class $C$ (w.r.t $L$) with size $d$ if when given ordered samples $S = \{x_1, \cdots, x_m\}$, $A$ outputs $B \subseteq S$ s.t.
1. $|B| \leq d$.
2. $hyp(B)$ is consistent with $S$.

• **Theorem 3:** A compression scheme is a PAC learner for class $C$ when using a sample of size
$$m \geq \frac{2}{\epsilon} \ln \frac{1}{\delta} + 2d + \left(\frac{2d}{\epsilon}\right) \ln \frac{2}{\epsilon}.$$