CS761 Spring 2015 Homework 3

Assigned Feb. 22, Due Mar. 2, 2015 before class

Instructions:

• Homeworks are to be done individually.

• Typeset your homework in latex using this file as template (e.g. use pdflatex). We do not accept hand-written homeworks. Show your derivations.

• Hand in a single-sided printed copy of your homework before class. Homework will no longer be accepted once the lecture starts.

• Unless explicitly specified in the questions, you do not need to hand in any code.

• Fill in your name and email below. This will produce a page separated from the rest of your homework. We will do “double blind review.” Do not include identifying information (e.g. your name) in the rest of your homework.

Name:
Email:
1. Let $\mathcal{H}$ be some hypothesis class. For any $h \in \mathcal{H}$, let $|h|$ be the description length of $h$ according to some fixed description language. Consider the Minimum Description Length learning algorithm:

$$h_{SRM} \in \arg\min_{h \in \mathcal{H}} \mathcal{L}(h) + \sqrt{\frac{|h| + \log(2/\delta)}{2n}}.$$ 

For any $B > 0$, let $\mathcal{H}_B = \{h \in \mathcal{H} : |h| \leq B\}$, and define

$$h_B^* = \arg\min_{h \in \mathcal{H}_B} \mathcal{L}(h)$$

as the reference classifier. Provide a bound on $\mathcal{L}(h_{SRM}) - \mathcal{L}(h_B^*)$ in terms of $B$, the confidence parameter $\delta$, and the training set size $n$.

2. (a) Consider a hypothesis class $\mathcal{H} = \bigcup_{i=1}^{\infty} \mathcal{H}_i$ where for every $i$, $\mathcal{H}_i$ is finite. Find a weighting function $w : \mathcal{H} \mapsto [0,1]$ such that $\sum_{h \in \mathcal{H}} w(h) \leq 1$, and such that for all $h \in \mathcal{H}$, $w(h)$ is determined by $i(h) = \min\{i : h \in \mathcal{H}_i\}$ and by $|\mathcal{H}_i|$.

(b) Define such a function $w$ when for all $i$, $\mathcal{H}_i$ is countable infinite.

3. This question relates Rademacher complexity to the VC dimension. For any finite $A \in \mathbb{R}^n$, define $b = \sup_{a \in A} \left(\sum_{i=1}^{n} a_i^2\right)^{1/2}$. The following theorem bounds the Rademacher complexity

$$R(A) = \frac{1}{n} \mathbb{E}_{\sigma} \left(\sup_{a \in A} \sum_{i=1}^{n} \sigma_i a_i\right) \leq \frac{b \sqrt{2 \log |A|}}{n}.$$ 

Fix a training sample $S$ of size $n$, some (potentially infinite) hypothesis class $\mathcal{H}$ with $\text{VCdim}(\mathcal{H}) > 2$, and loss function $\ell(h, x, y) \in [-1, 1]$. Use the theorem above, prove that

$$R(\ell \circ \mathcal{H} \circ S) \leq \sqrt{\frac{2d \log n}{n}}.$$ 

4. This question asks you to be creative. As we have seen in class, a hypothesis class $\mathcal{H}$ can be defined over physical or psychological entities such as all the rules humans can come up with to classify words.

(a) Suggest another $\mathcal{H}$ that has a clear real-world interpretation. Be sure to define $\mathcal{X}, \mathcal{Y}$ (if appropriate), the loss function $\ell()$, whether there is a training set $S$, a learning algorithm $A$, and if it makes sense to define $\mathcal{L}_D(A(S))$.

(b) Suggest a method to estimate the capacity of your $\mathcal{H}$.

As a last resort, someone can probably blame Punxsutawney Phil for this cold weather...