HW# 2

Entrega: segunda-feira, 12/10/2015 (23:59)
Atenção: Justifique todas suas respostas.

1. (**Problem 2.1**) In $E_{\text{out}}(g) \leq E_{\text{in}}(g) + \sqrt{\dfrac{1}{2N} \ln \dfrac{2M}{\delta}}$, set $\delta = 0.03$ and let

$$\epsilon(M, N, \delta) = \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}}$$

.

   (a) For M = 1, how many examples do we need to make $\epsilon \leq 0.05$?
   (b) For M = 100, how many examples do we need to make $\epsilon \leq 0.05$?
   (c) For M = 10000, how many examples do we need to make $\epsilon \leq 0.05$?

2. (**Problem 2.5**) prove by induction that $\sum_{i=0}^{D} \binom{N}{i} < N^D + 1$, hence

$$m_{\mathcal{H}}(N) \leq N^{d_{\text{VC}}} + 1$$

.

3. **Problem 2.14** Let $\mathcal{H}_1, \mathcal{H}_2, ..., \mathcal{H}_k$ be K hypothesis sets with finite VC dimension d_{VC}. Let $\mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2 \cup ... \cup \mathcal{H}_k$ be the union of these models.

   (a) Show that $d_{\text{VC}}(\mathcal{H}) < K(d_{\text{VC}} + 1)$.
   (b) Suppose that ℓ satisfies $2^\ell < 2K\ell^{d_{\text{VC}}}$. Show that $d_{\text{VC}}(\mathcal{H}) \leq \ell$.
   (c) Hence, show that

$$d_{\text{VC}}(\mathcal{H}) \leq \min(K(d_{\text{VC}} + 1), 7(d_{\text{VC}} + K) \log_2(d_{\text{VC}} K)).$$

   That is, $d_{\text{VC}}(\mathcal{H}) =)(\max(d_{\text{VC}}, K) \log_2 \max(d_{\text{VC}}, K))$ is not too bad.

4. **Problem 2.20** There are a number of bounds on the generalization error ϵ, all holding with probability at least $1 - \delta$.

   (a) Original VC-bound:

$$\epsilon \leq \sqrt{\frac{8}{N} \ln \frac{4m_{\mathcal{H}}(2N)}{\delta}}$$

   (b) Rademacher penalty Bound:

$$\epsilon \leq \sqrt{\frac{2\ln(2Nm_{\mathcal{H}}(N))}{N}} + \sqrt{\frac{2}{N} \ln \frac{1}{\delta}} + \frac{1}{N}$$

(c) Parrondo and Van den Broek:

$$\epsilon \leq \sqrt{\frac{1}{N}\left(2\epsilon + \ln\frac{6m_{\mathcal{H}}(2N)}{\delta}\right)}$$

(d) Devroye:

$$\epsilon \leq \sqrt{\frac{1}{2N}\left(4\epsilon(1+\epsilon) + \ln\frac{4m_{\mathcal{H}}(N^2)}{\delta}\right)}$$

Note that (c) and (d) are implicit bounds in ϵ. Fix $d_{\text{VC}} = 50$ and $\delta = 0.05$ and plot these bounds as a function of N. Which is best?

5. (**Problem 2.22**) When there is noise in the data, $E_{\text{out}}(g) = \mathbb{E}_{x,y}[(g(x) - y(x))^2]$, where $y(x) = f(x) + \epsilon$. If ϵ is a zero mean noise random variable with variance σ^2, show that the bias variance decomposition becomes

$$\mathbb{E}_{\mathcal{D}}[E_{\text{out}}(g)] = \sigma^2 + \text{bias} + \text{var}.$$