Exercises

Machine Learning: Foundations and Applications
MATH 260J

Paul J. Atzberger
http://atzberger.org/

1. Consider a game where we see m coin flips and we need to guess which of two coins A or B generated the data. Consider the case when the coin A has heads probability $p_A = 1/2 + \gamma$ and the coin B has heads probability $p_B = 1/2 - \gamma$. Use $\gamma = 0.1$. Suppose we use the following strategy of selecting the coin $h \in \{A, B\}$: (i) select A if the m flips had more heads, (ii) select B if the m flips had more tails.

   At most how many coin tosses m do we need to observe so that our strategy would identify the correct coin 99% of the time?

   Hint: Use Hoeffding's Inequality to get a lower bound on m so that
   $\Pr[|\frac{1}{m}S_m^{(i)} - p_i| \geq t] \leq 2\exp(-2t^2 m) < \delta = 0.01$, where $i \in \{A, B\}$.

2. Consider the concept class of concentric circles of the form $\{(x, y) \mid x^2 + y^2 \leq r^2\}$ for $r > 0,\ r \in \mathbb{R}$. Show this is (ϵ, δ)-PAC-learnable from a training data set of size $m \geq (1/\epsilon)\log(1/\delta)$.

3. VC-Dimension: Determine the VC-Dim($\mathcal{H}$) of each of the following hypothesis spaces:

   (a) Classifiers using polynomials of degree n, $\mathcal{H} = \{h \mid h(x) = \text{sign}(p(x)),\ p \in \mathbb{P}^n\}$.

   (b) For a finite set $\mathcal{X}$ and number $k \leq |\mathcal{X}|$, let $\mathcal{H}_k = \{h \in \{0,1\}^{\mathcal{X}} \mid\ |\{x \mid h(x) = 1\}| = k\}$ (i.e. all functions that can assign only exactly k points in $\mathcal{X}$ the label 1).

   (c) For a finite set $\mathcal{X}$ and number $k \leq |\mathcal{X}|$, let $\mathcal{H}_{\leq k} = \{h \in \{0,1\}^{\mathcal{X}} \mid\ |\{x \mid h(x) = 1\}| \leq k\}$ (i.e. all functions that can assign at most k points in $\mathcal{X}$ the label 1).

4. Consider a random variable X that is non-negative satisfying the inequality $\Pr[X > t] \leq c\exp(-2mt^2)$ for all $t > 0$. Show that $E[X^2] \leq \log(ce)/2m$.

   Hints: Do this by using that $E[X^2] = \int_0^\infty \Pr[X^2 > t]dt = \int_0^u \Pr[X^2 > t]dt + \int_u^\infty \Pr[X^2 > t]dt$ for any choice of $u > 0$. For the first term, use that probabilities are always bounded by one. Optimize the obtained bound over u.

5. Consider a family of functions $f^{(m)} : \mathcal{X}^m \to \mathbb{R}$ on a sample space $\mathcal{X}$ and a sequence c_i with $\sum_{i=1}^\infty c_i^2 < \infty$. Suppose that $f^{(m)}$ has bounded dependence on parameters in the sense

   $$|f^{(m)}(x_1, \ldots, x_i, \ldots, x_m) - f^{(m)}(x_1, \ldots, x_i^*, \ldots, x_m)| \leq c_i. \tag{1}$$

   For short-hand we denote $f(s) = f^{(m)}(x_1, \ldots, x_i, \ldots, x_m)$.

   Consider the case when $f^{(m)} = (1/m)\sum_{k=1}^m X_k$ for i.i.d random variables $X_i \in \mathcal{X}$ with $|X_i| \leq C$. Show this has bounded dependence.

   How many samples m do we need so that the values $f(S)$ and its mean value $E[f(S)]$ are within the distance 0.1 and this occurs 99% of the time?

In other words, establish the following bound and find for what m we have

$$\Pr[|f(S) - E[f(S)]| \geq \epsilon] \leq 2 \exp\left(-2\epsilon^2 / \sum_{i=1}^{m} c_i^2\right) < \delta, \tag{2}$$

where $\delta = 0.01$ and $\epsilon = 0.1$. Hint: Use McDiarmids Inequality with $c_i = C/i$.