Introduction to Machine Learning, Fall 2012

Problem Set 2: Machine learning theory

Due: Monday, October 8, 2012 by 4pm (hand in to Hong Tam in the reception area of the 12th floor, 715 Broadway)

Important: *See problem set policy on the course web site.* You must show **all** of your work and be rigorous in your writeups to obtain full credit.

---

1. (15 points) Probability:

   (a) Prove that if X and Y are independent real-valued random variables, then $E[XY] = E[X]E[Y]$. Also, prove that this equality need not hold when X and Y are not independent (i.e., give an example). You may assume that X and Y are discrete variables having a finite set of possible values.

   (b) Suppose one card is picked at random from each of four ordinary decks of cards (i.e., each deck has 52 cards). What is the probability that an ace of spades is among the four selected cards? Answer this question with an *approximate* upper bound using the union bound, and also give an *exact* answer. Evaluate your answers numerically.

   Hint for the exact answer: First compute the probability that *none* of the four selected cards are ace of spades.

   (c) Suppose now that all four decks of cards are shuffled together and a set of four cards are randomly picked from this huge deck. Now what is the probability that an ace of spades is among the four selected cards? Give an exact answer. Can the union bound be used also in this case? If so, what approximate answer does it give? If not, why not? As before, evaluate your answers numerically.

   Hint for the exact answer: Let X_i be the event that the i'th card is not an ace of spades. The conditional probability of event A given B is defined as $\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$. The chain rule of probability states that $\Pr(X_1 \cap X_2 \cap X_3 \cap X_4) = \Pr(X_1) \Pr(X_2 \mid X_1) \Pr(X_3 \mid X_1, X_2) \Pr(X_4 \mid X_1, X_2, X_3)$.

2. (15 points) VC-dimension:

   (a) Show that the VC-dimension of a finite hypothesis set H is at most $\log_2 |H|$.

   (b) Show that the VC-dimension of the set of all closed balls in $\mathbb{R}^d$, that is sets of the form $\{\vec{x} \in \mathbb{R}^d : ||\vec{x} - \vec{x}_0||_2 \leq r\}$ for some $\vec{x}_0 \in \mathbb{R}^d$ and $r \geq 0$, is at most $d + 2$.

   Hint: Recall that the VC-dimension of the set of linear classifiers $\vec{w} \cdot \vec{x} \geq b$ in dimension n (i.e., $\vec{w} \in \mathbb{R}^n$) is $n + 1$. Construct a feature mapping to reduce this hypothesis class to a subset of linear classifiers of dimension $d + 1$, and then apply this result.

   (c) Consider the hypothesis class $\mathcal{H}_\alpha$ defined on the real line $x \in \mathbb{R}$ and parameterized by a single parameter α, given by $\mathcal{H}_\alpha = \{x : x \in [\alpha, \alpha + 1] \text{ or } x \in [\alpha + 2, +\infty]\}$. Show that the VC-dimension of $\mathcal{H}_\alpha$ is exactly 3.
   (Recall that to prove that the VC-dimension is 3, you must (i) demonstrate a set of three points that are shattered by the hypothesis class, and (ii) demonstrate that any set of four or more points *cannot* be shattered by the hypothesis class.)

Acknowledgements: Problem 1 is adapted from an assignment developed by Rob Schapire at Princeton. Problem 2 is adapted from an assignment by Mehryar Mohri.