Homework 2

Machine Learning: Foundations and Applications — Paul J. Atzberger
MATH CS 120 — http://atzberger.org/

1. 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.

2. Consider a game where we see coin flips and need to decide which of two coins A and B generated the data. Consider the case when the coins have probabilities of heads $p_A = 1/2+\gamma$ and $p_B = 1/2 - \gamma$ with $\gamma = 0.1$. Suppose we use the strategy of attributing the coin based on a sample of m flips if we saw that most were heads or most were 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 an upper 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 $i \in \{A, B\}$.

3. 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 $c_i = C/i$.

4. Consider k Nearest Neighbor (k-NN) classifiers. Suppose the input data space has features from the unit cube in d-dimensional space and there are two classes we want to distinguish. Suppose that in order to capture well the classes, we need a prototype within a distance at most ϵ of any given input $x \in [0,1]^d$. Give an estimate of the number m of training samples (prototypes) needed to ensure this distance requirement holds. Consider here the case of the Euclidean distance. How does the number m of prototypes scale with dimension d? How many samples m do you need when $\epsilon = 10^{-1}$ and $d = 100$ if you use the Euclidean distance?

5. Suppose for a data point x_0 in d dimensional space the conditional probability of a neighboring data point X is distributed uniformly within the unit sphere. Compute the probability density

of $\rho(r)$ where $\Pr\{r_1 \leq |X - x_0| \leq r_2\} = \int_{r1}^{r2} \rho(r)dr$. Show as $d \to \infty$ for any $\epsilon > 0$ that $\Pr\{1 - \epsilon \leq |X - x_0| \leq 1\} \to 1$. Give an upper bound on $|\Pr\{1 - \epsilon \leq |X - x_0| \leq 1\} - 1|$ in terms of ϵ and d. This result shows that when d corresponds to a high dimensional space we have that the neighboring data points tend to distribute near to the surface of the sphere. For $d = 100$ what is the probability that the neighbor X for x_0 is within the distance $r = 10^{-1}$? For $\epsilon = 10^{-1}$ how large must d be for $\Pr\{1 - \epsilon \leq |X - x_0| \leq 1\} = 99\%$? Explain briefly what implications this might have for k-NN and other methods.