Problem 4.1
Let $\mathcal{A}$ be a finite class of sets (i.e., $|\mathcal{A}| < \infty$). Determine upper bounds on the shatter coefficients and VC dimension of $\mathcal{A}$. Provide an example for which your upper bounds are tight.

Problem 4.2
Determine the VC dimension of the following classes of sets:

(a) The class of sets in $\mathbb{R}^d$:
$$\mathcal{A} = \{(-\infty, a_1] \times (-\infty, a_2] \times \ldots \times (-\infty, a_d] \mid (a_1, \ldots, a_d) \in \mathbb{R}^d\}.$$ 

(b) The class of sets in $\mathbb{R}^d$:
$$\mathcal{A} = \{(b_1, a_1] \times (b_2, a_2] \times \ldots \times (b_d, a_d] \mid (a_1, \ldots, a_d), (b_1, \ldots, b_d) \in \mathbb{R}^d\}.$$ 

(c) The class of all closed balls in $\mathbb{R}^2$—that is, $\mathcal{A}$ is the class of all subsets of the form
$$\{x \in \mathbb{R}^2 \mid \sum_{i=1}^{2}(x_i - a_i)^2 \leq R, \text{ for some } (a_1, a_2) \in \mathbb{R}^2 \text{ and } R > 0\}.$$ 

Problem 4.3
Define the binary entropy function $h : [0, 1] \to [0, 1]$ by $h(t) = -t \log t - (1-t) \log(1-t)$ (in base $\log e$). In this problem, we prove that for all $k \leq n/2$,
$$\sum_{i=0}^{k} \binom{n}{i} \leq \exp \left( n h \left( \frac{k}{n} \right) \right).$$ 

(a) Show that $\sum_{i=0}^{k} \binom{n}{i} = 2^n \mathbb{P}[Z \geq n - k]$, where $Z$ is a binomial $(n, 1/2)$ variate.

(b) Use the Chernoff bounding technique to derive a sharp upper bound on the binomial tail probability. (Hint: It is not sufficient to use the sub-Gaussian tail bound from the boundedness of Bernoulli variables.)

Problem 4.4
A sequence $V_1, V_2, \ldots$ of integrable random variables form a martingale difference sequence (MDS) with respect to another sequence $Z_1, Z_2, \ldots$ if
$$\mathbb{E}[V_{i+1} \mid Z_1, \ldots, Z_i] = 0 \quad \text{for all } i = 1, 2, \ldots.$$
(a) Given a function $f(Z_1, \ldots, Z_n)$ and a sequence of i.i.d. random variables $Z_1, \ldots, Z_n$, define $V = f(Z_1, \ldots, Z_n) - \mathbb{E}[f(Z_1, \ldots, Z_n)]$ and consider the sequence given by
\[ V_1 = \mathbb{E}[V \mid Z_1], \quad \text{and} \quad V_k = \mathbb{E}[V \mid Z_1, \ldots, Z_k] - \mathbb{E}[V \mid Z_1, \ldots, Z_{k-1}] \quad \text{for } k = 2, 3, \ldots, n. \]
Show that \{\{V_i\}\} is an MDS with respect to \{Z_i\}.

(b) Show that $\text{var}(V) = \sum_{i=1}^{n} \text{var}(V_i)$, which yields a useful decomposition for bounding the variance of $V$.

(c) Suppose that $f$ satisfies the bounded difference property with parameter $c_i$ for each $i = 1, \ldots, n$. Show that $\text{var}(V) \leq \frac{1}{4} \sum_{i=1}^{n} c_i^2$.

Problem 4.5
(Clustering) Let $X^{(1)}, \ldots, X^{(n)}$ be i.i.d. random variables from a distribution supported on $[-B, B]^d$ for some $B < +\infty$. The k-means method of clustering or vector quantization is based on choosing a set of $k$ vectors $a^1, \ldots, a^k$ in $\mathbb{R}^d$ (representing centers of $k$ clusters) to minimize the empirical squared error:
\[ \hat{M}_n(a^1, \ldots, a^k) = \frac{1}{n} \sum_{i=1}^{n} \min_j \|X^{(i)} - a^j\|_2^2 \]

The population error of the clustering can be measured by the quantity
\[ M(a^1, \ldots, a^k) = \mathbb{E} \left[ \min_j \|X - a^j\|_2^2 \mid X^{(1)}, \ldots, X^{(n)} \right] \]
where $X$ is an independent draw from the same distribution.

(a) If $(a^1, \ldots, a^k)$ are a set of empirically optimal cluster centers, show that
\[ M(a^1, \ldots, a^k) = \inf_{b^1, \ldots, b^k \in \mathbb{R}^d} M(b^1, \ldots, b^k) \leq 2 \sup_{b^1, \ldots, b^k \in \mathbb{R}^d} |\hat{M}_n(b^1, \ldots, b^k) - M(b^1, \ldots, b^k)|. \]

(b) Show that for all $\epsilon > 0$ with $n \epsilon^2 \geq 2$,
\[ \mathbb{P} \{ \sup_{b^1, \ldots, b^k \in \mathbb{R}^d} |\hat{M}_n(b^1, \ldots, b^k) - M(b^1, \ldots, b^k)| > \epsilon \} \leq C n^{2k(d+1)} \exp \left( - \frac{n \epsilon^2}{32B^4} \right), \]
for some constant $C$ independent of $(n, k, d, B)$. (*Hint: The reasoning in problem 4.2(c) could be relevant for part of your argument.*)

(c) Conclude that the population error of the empirically optimal clustering converges in probability to $\inf_{b^1, \ldots, b^k \in \mathbb{R}^d} M(b^1, \ldots, b^k)$ as $n \to +\infty$.

Problem 4.6
Let $X^{(1)}, \ldots, X^{(n)}$ be i.i.d. random variables in $\mathbb{R}$ with probability distribution $\mathbb{P}$. For some class of subsets $\mathcal{A}$, define $Z_n = \sup_{A \in \mathcal{A}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}[X^{(i)} \in A] - \mathbb{P}[A] \right\}$. Show that $Z_n \xrightarrow{p} 0$ implies that $Z_n \xrightarrow{a.s.} 0$. (*Hint: The bounded difference inequality from class and the Borel-Cantelli lemma could be useful. Recall Borel-Cantelli: if for each fixed $\epsilon > 0$, we have $\sum_{n=1}^{\infty} \mathbb{P}[|Z_n| > \epsilon] < +\infty$, then $Z_n \xrightarrow{a.s.} 0$.)