Please provide complete and well-written solutions to the following exercises.

Due November 8, at the beginning of class.

Homework 3

Exercise 1. Let $a > 0$. Let $X^{(1)}, \ldots, X^{(k)} \in \mathbf{R}^n$ be independent identically distributed samples from a Gaussian random vector with mean $(a, 0, \ldots, 0)$ and identity covariance matrix). Let $X^{(k+1)}, X^{(k+2)}, \ldots, X^{(2k)} \in \mathbf{R}^n$ be independent identically distributed samples from a Gaussian random vector with mean $(-a, 0, \ldots, 0)$, where $a > 0$ is known. Define $y_1 = \cdots = y_k := 1$, and $y_{k+1} = \cdots = y_{2k} := -1$. With this "planted" data with $a = 5, n = 3, k = 20$, run the perceptron algorithm. (You will have to modify the algorithm to terminate in case a separating hyperplane does not exist). Then, run a support-vector machine, and compare the quality of the results and the run-time.

Exercise 2. Show that any boolean function $f\colon \{0,1\}^n \to \{0,1\}$ can be written as a DNF formula of size at most $n2^n$. (Here the size of the DNF formula f refers to the number of AND and OR operations that are used to construct f.) (Hint: consider the function $f\colon \{0,1\}^n \to \{0,1\}$ defined by $f := 1_{(0,\ldots,0)}$. This function can be written as a conjunction with $n-1$ AND operations.)

Exercise 3. Give an *efficient* PAC learning algorithm for the class of axis-aligned rectangles in the plane. (So, in this case, $\mathcal{F}$ is the function class consisting of all functions of the form 1_R where $R \subseteq \mathbf{R}^2$ is a rectangle whose edges are parallel to the x and y axes, respectively.) (Note: if you want you can compute the VC-dimension of this class, but simply appealing to a VC-dimension bound does not give an efficient algorithm a priori.) (Hint: consider the smallest rectangle containing all points that are known, via random sampling, to be contained in the unknown rectangle.)

(Optional: can you generalize your results to axis-aligned boxes in $\mathbf{R}^n$?)

Exercise 4.

- Give an example of a class of boolean functions $\mathcal{F}$ such that the VC-dimension of $\mathcal{F}$ is 1, while $\mathcal{F}$ contains infinitely many functions.
- Give an example of a class of boolean functions $\mathcal{F}$ such that the VC-dimension of $\mathcal{F}$ is not finite.
- For any $I \subseteq \{1, \ldots, n\}$, define the parity function $h_I\colon \{0,1\}^n \to \{0,1\}$ by
$$h_I(x) := \Big(\sum_{i \in I} x_i\Big) \mod 2, \qquad \forall\, x = (x_1, \ldots, x_n) \in \{0,1\}^n.$$
Find the VC-dimension of the class $\mathcal{F} = \{h_I \colon I \subseteq \{1, \ldots, n\}\}$ of all parity functions.

Exercise 5. Suppose $x, a \geq 1$. Assume that
$$x < a \log(x).$$

Then

$$x < 2a \log a.$$

Exercise 6. High-dimensional geometry is much different than low-dimensional geometry, as this exercise demonstrates.

- Show that "most" of the mass of an n-dimensional Gaussian is concentrated on the sphere of radius $\sqrt{n}$ centered at the origin. That is, if $X_1, \ldots, X_n$ are n i.i.d. standard Gaussian random variables, then

  $$\lim_{n\to\infty} \mathbb{P}(X_1^2 + \cdots + X_n^2 \in (n - \sqrt{3n}, n + \sqrt{3n}) \geq 2/3.$$

  In fact, you should be able to compute the limit exactly.
- Generally, "most" of the mass of a high-dimensional convex body is concentrated near the surface of the body. Let Vol_n denote the usual volume in $\mathbf{R}^n$ (so that the volume of a unit square $[0,1]^n$ is 1.) For example, show that, for any $\varepsilon > 0$,

  $$\lim_{n\to\infty} \mathrm{Vol}_n\Big([-\frac{1}{2}(1-\varepsilon), \frac{1}{2}(1-\varepsilon)]^n\Big) = 0.$$
- Let $B_n := \{x \in \mathbf{R}^n \colon\ ||x|| \leq 1\}$ be the unit ball centered at the origin. Show that

  $$\lim_{n\to\infty} \mathrm{Vol}_n(B_n) = 0.$$
- Let $C_n = \{x \in \{[-1/2, 1/2]^n \colon \exists\, y \in \{-1/2, 1/2\}^n \text{ such that } ||x - y|| \leq 1/2\}\}$ be the union of balls of radius $1/2$ centered at the corners of the hypercube $[-1/2, 1/2]^n$. Let $D_n := \{x \in \mathbf{R}^n \colon\ ||x|| \leq r\}$ be a ball of radius r centered at the origin, where r is chosen to be as large as possible so that D_n does not intersect the interior of C_n. (Put another way, D_n is tangent to the balls C_n.) Find

  $$\lim_{n\to\infty} \mathrm{Vol}_n(D_n).$$

  Before you do the computation, try to guess what the answer should be.

Exercise 7. Let X be a real-valued random variable with mean zero. Then the following are equivalent

- $\exists\, a > 0$ such that, for all $t \in \mathbf{R}$, $\mathbb{E}e^{tX} \leq e^{t^2a^2/2}$.
- $\exists\, b > 0$ such that, for all $t > 0$, $\mathbb{P}(|X| > t) \leq 2e^{-bt^2}$.
- $\exists\, c > 0$ such that $\mathbb{E}e^{cX^2} \leq 2$.
- $\exists\, d > 0$ such that $(\mathbb{E}\,|X|^p)^{1/p} \leq d\sqrt{p}$, $\forall\, p \geq 1$.

(If you need hints look at Proposition 2.5.2 in Vershynin's book.)

Exercise 8. Show that $||\cdot||_{\psi_2}$ is a norm on the set of sub-Gaussian random variables.

Exercise 9. Let $Y_1, Y_2, \ldots$ be a sequence of sub-Gaussian random variables. (These random variables are not assumed to be independent.) Prove that

$$\mathbb{E} \max_{i\geq 1} \frac{|Y_i|}{\sqrt{1 + \log(i+1)}} \leq 100 \sup_{i\geq 1} ||Y_i||_{\psi_2}.$$

Conclude that, for any integer $n \geq 2$, we have

$$\mathbb{E} \max_{1 \leq i \leq n} |Y_i| \leq 100\sqrt{\log n} \cdot \max_{1 \leq i \leq n} ||Y_i||_{\psi_2}.$$

(Hint: there are a few related ways to do this. Your first step could use the union bound of the form $\mathbb{P}(\max_{i \geq 1} X_i > t) \leq \sum_{i \geq 1} \mathbb{P}(X_i > t)$.)

Exercise 10. Using the argument for Dudley's inequality, deduce the following concentration inequality. For any $u > 0$,

$$\mathbb{P}\Big(\sup_{a \in A} X_a \leq 10^3 c \int_0^\infty \sqrt{\log \mathcal{N}(A, d, \varepsilon)} d\varepsilon + u \cdot \mathrm{diam}(A)\Big) \geq 1 - 2e^{-u^2}.$$

Here $\mathrm{diam}(A) := \sup_{a,a' \in A} d(a, a')$. (Hint: show $\sup_{a \in A} \left|X_{b_{k+1}(a)} - X_{b_k(a)}\right| \leq 2^{-k}\sqrt{\log |\mathcal{N}_{k+1}|} + u_k$ with probability at least $1 - 2e^{-u_k^2}$. Then sum over k, use the union bound, and choose the u_k appropriately, e.g. try $u_k = u + \sqrt{k - m}$.)

Exercise 11 (Optional)**.** Show that Empirical Risk Minimization of Linear Threshold Functions in $\mathbf{R}^n$ is computationally hard. More precisely, we consider the sequence of problems in which the dimension n is large, c is a fixed positive integer, and the number of samples m is equal to cn.

Hint: You can prove the hardness by a reduction from the following problem, known as MAX FS: We are given a system of linear inequalities, $Ax > b$ with A an $m \times n$ matrix, $b \in \mathbf{R}^m$. That is, we are given a set of m linear inequalities in n variables $(x_1, \ldots, x_n) =: x$. The goal is to find the largest possible subset of $\{1, \ldots, m\}$ that has a solution (i.e. a feasible solution). That is, the goal is to find

$$\max\{|S| : S \subseteq \{1, \ldots, m\}, \exists\, x \in \mathbf{R}^n, (Ax)_i > b_i, \quad \forall\, i \in S\}.$$

It has been shown (Sankaran 1993) that the problem MAX FS is NP-hard. That is, if one could solve MAX FS in polynomial time (in n), then P=NP.

Show that any algorithm that minimizes the empirical risk for any training sample $S \in (\mathbf{R}^n \times \{-1, 1\})^m$ can be used to solve the MAX FS problem with parameters m, n. Hint: Define a mapping that transforms linear inequalities in n variables into labeled points in $\mathbf{R}^n$, and a mapping that transforms vectors in $\mathbf{R}^n$ to halfspaces, such that a vector w satisfies an inequality q if and only if the labeled point that corresponds to q is classified correctly by the halfspace corresponding to w. Conclude that the problem of empirical risk minimization for halfspaces in also NP-hard (that is, if it can be solved in time polynomial in the sample size, m, and the Euclidean dimension, n, then P=NP).