Exercises for CSCI5010

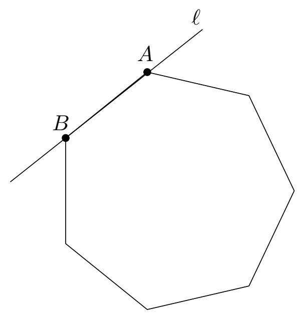
Prepared by Yufei Tao

Problem 1. Define $\mathcal{R}$ as the set of all halfplanes in $\mathbb{R}^2$. Prove: The VC-dimension of the range space $(\mathbb{R}^2, \mathcal{R})$ is 3.

Problem 2. Let $\mathcal{R}$ be the set of all triangles in $\mathbb{R}^2$. Let S be an arbitrary set of lines in $\mathbb{R}^2$. Given any triangle $q \in \mathcal{R}$, define $q(S)$ to be the set of lines in S that intersect q. In addition define

$$\mathcal{R}_S \quad = \quad \{q(S) \mid q \in \mathcal{R}\}.$$

Let G be a regular 7-gon. Let S be the set of support lines of the 7 edges of G (the support line ℓ of the edge $\overline{AB}$ is shown below). Prove: $\mathcal{R}_S = 2^S$.



Problem 3. Let $(X, \mathcal{F})$ be a set system with VC-dimension λ. Fix an arbitrary element $e \in X$. Define

$$\begin{aligned} X' &= X \setminus \{e\} \\ \mathcal{F}'' &= \{Q \setminus \{e\} \mid Q \in \mathcal{F}, e \in Q \text{ and } Q \setminus \{e\} \notin \mathcal{F}\} \\ \mathcal{F}''' &= \{Q \mid Q \in \mathcal{F} \text{ and } e \notin Q\} \\ \mathcal{F}_2 &= \mathcal{F}'' \cup \mathcal{F}''' \end{aligned}$$

Prove: the VC-dimension of the set system $(X', \mathcal{F}_2)$ is at most λ.

Remark: The above notations follow those used in our lecture videos. This problem should be fairly easy, and its solution completes our proof of Sauer's lemma.

The next few problems are designed to help you understand how Sauer's lemma can be used to prove the existence of a small ϵ-net. In all the following problems, we will fix ϵ to any value satisfying $0 < \epsilon < 1$.

Problem 4. Let $(X, \mathcal{F})$ be any set system where X is a finite set of size n. Let Q be an arbitrary set in $\mathcal{F}$ with size at least ϵn. Suppose that we create a sample set $S \subseteq X$ by repeating the following operation $s \geq 1$ times on an initially empty S: take an element e from X uniformly at random and add e to S. Note that the final size of S may be less than s because it is possible for the same element to be taken more than once. Prove: $\mathbf{Pr}[S \cap Q = \emptyset] \leq 1/e^{\epsilon \cdot s}$.

Hint: As long as you understand the problem, you should find it fairly easy, but you will need the inequality $1 + x \le e^x$ for any real value x.

Problem 5. Let $(X, \mathcal{F})$ be any set system where X is a finite set of size n. Denote by λ the VC-dimension of $(X, \mathcal{F})$. Given a parameter $s \ge 1$, take a sample set S of X in the way described in Problem 4. Let B be the *bad event* where at least one $Q \in \mathcal{F}$ satisfies both conditions below:

- $|Q| \ge \epsilon n$;
- $Q \cap S = \emptyset$.

Prove: $\mathbf{Pr}[B] \le \Phi_\lambda(n)/e^{\epsilon \cdot s}$.

Hint: There is a rudimentary tool called the *union bound* in probability theory:

> **Union bound.** Let $A_1, A_2, ..., A_k$ be arbitrary events where $k \ge 2$ (these events can be correlated to each other). Let $\cup_{i=1}^k A_i$ denote the event that at least one of $A_1, A_2, ..., A_k$ occurs. It holds that
>
> $$\mathbf{Pr}[\cup_{i=1}^k A_i] \le \sum_{i=1}^k \mathbf{Pr}[A_i].$$

Problem 6. Let $(X, \mathcal{F})$ be any set system where X is a finite set of size n. Denote by λ the VC-dimension of $(X, \mathcal{F})$. Set $s = \lceil c \cdot \frac{\lambda}{\epsilon} \log_2 n \rceil$ where c is a constant to be determined later. Given this value of s, take a sample set S of X in the way described in Problem 4. Prove: when c is larger than a certain constant (which does not depend on n, ϵ, and λ), the probability for S to be an ϵ-net of $(X, \mathcal{F})$ is at least $1 - 1/n^2$.

Hint: It suffices to use the trivial fact $\Phi_\lambda(n) = \sum_{i=0}^{\lambda} \binom{n}{i} < 1 + \lambda \cdot n^\lambda$ for any $\lambda \in [0, n]$. This inequality is very loose. The next statement gives a tighter result: for $\lambda \ge 2$, it holds that $\Phi_\lambda(n) \le n^\lambda$.

Problem 7. Let $(X, \mathcal{F})$ be any set system where X is a finite set of size n. Denote by λ the VC-dimension of $(X, \mathcal{F})$. Prove: there is an ϵ-net of $(X, \mathcal{F})$ with size $O(\frac{\lambda}{\epsilon} \log n)$.

Hint: Obviously, you need the result of Problem 6. At first glance, Problem 7 appears to follow from Problem 6 immediately. Well, not quite — the sampling process in Problem 6 may fail to yield an ϵ-net with a positive probability.