Advanced Topics in Artificial Intelligence: EE6180
Indian Institute of Technology Madras — September 10, 2019
Instructor: Abhishek Sinha

Mid Term

- The Mid-Term is due on **Friday, September 20, 2019** in the class.
- Each problem carries 10 points.
- Collaboration among the students is strictly prohibited.

---

1. **(Concentration and kernel density estimation)** Let $\{X_i\}_{i=1}^n$ be an i.i.d. sequence of random variables drawn from a density f on the real line. A standard estimate of f is the *kernel density estimate*:

$$\hat{f}_n(x) \equiv \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right),$$

where $K : \mathbb{R} \to [0, \infty)$ is a kernel function satisfying $\int_{-\infty}^{\infty} K(t)dt = 1$, and $h > 0$ is a bandwidth parameter. Suppose that we assess the quality of $\hat{f}_n$ using the L^1-norm $||\hat{f}_n - f||_1 \equiv \int_{-\infty}^{\infty} |\hat{f}_n(t) - f(t)|dt$. Prove that

$$\mathbb{P}\left(||\hat{f}_n - f||_1 \geq \mathbb{E}||\hat{f}_n - f||_1 + \delta\right) \leq \exp(-n\delta^2/c),$$

for some absolute constant $c > 0$.

2. **(VC dimension and No-Free-Lunch)** Let $\mathcal{H}$ be a class of functions from $\mathcal{X}$ to $\{0, 1\}$.

   (a) Prove that if $\mathsf{VCdim}(\mathcal{H}) \geq d$, for any d, then for some probability distribution $\mathcal{D}$ over $\mathcal{X} \times \{0, 1\}$, for every sample size m,

$$\mathbb{E}_{S\in\mathcal{D}^m}[L_D(A(S))] \geq \min_h L_{\mathcal{D}}(h) + \frac{d-m}{2d}.$$

   (b) Hence show that for every $\mathcal{H}$ that is PAC learnable, $\mathsf{VCdim}(\mathcal{H}) < \infty$.

3. **(VC dimension of hyperplanes and hyperballs)** In this problem, we will compute the VC dimensions of d-dimensional hyperplanes (sets of the form $\boldsymbol{a}\cdot\boldsymbol{x} + b \leq 0, \boldsymbol{a} \in \mathbb{R}^d, b \in \mathbb{R}$) and hyperballs (sets of the form $||\boldsymbol{x} - \boldsymbol{c}||_2 \leq r, \boldsymbol{c} \in \mathbb{R}^d, r \geq 0$). For this, the following lemma from convex geometry will be useful.

Lemma 0.1 (Radon's Lemma) *Any set of d+2 points $S = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_{d+2}\} \subset \mathbb{R}^d$ can be partitioned into two disjoint sets A and B such that $\mathsf{conv}(A) \cap \mathsf{conv}(B) \neq \phi$, where $\mathsf{conv}(Z)$ denotes the convex hull of the set Z.*

We will now walk through a proof of Radon's Lemma. Arrange the points $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_{d+2}$ as the columns of a $d \times (d+2)$ matrix A. Append an all-one row at the end of the matrix A to obtain a $(d+1) \times (d+2)$ matrix $\tilde{A}$.

(a) Show that the columns of $\tilde{A}$ are linearly dependent, i.e., there exist reals $\alpha_1, \alpha_2, \dots, \alpha_{d+2}$, not all zero such that the following holds,

$$\begin{aligned} \sum_{i=1}^{d+2} \alpha_i \boldsymbol{x}_i &= \mathbf{0}, \\ \sum_{i=1}^{d+2} \alpha_i &= 0. \end{aligned} \tag{1}$$

(b) Divide the indices into two disjoint sets $I = \{i : \alpha_i \geq 0\}$ and $J = \{i : \alpha_i < 0\}$ and group the terms together in Eqn. (1). Conclude that there exists a point in the intersection of $\mathsf{conv}(x_i, i \in I)$ and $\mathsf{conv}(x_j, j \in J)$. This concludes the proof of Radon's lemma.

(c) Exhibit a set U of $d+1$ points in $\mathbb{R}^d$ which can be shattered by the hyperplanes. Hence VC dimension of the hyperplanes is at least $d+1$.

(d) With the help of Radon's lemma, show that given any set of $d+2$ points S in $\mathbb{R}^d$, any hyperplane can not shatter S. Hence VC dimension of the hyperplanes is strictly less than $d+2$.

(e) Hence conclude that the VC dimension of the class of d-dimensional hyperplanes is $d+1$.

(f) Repeat parts (c), (d), and (e) for the case of hyperballs.