Mathematics of Machine Learning
Homework 10 - Solutions

Instructors: Afonso S. Bandeira & Nikita Zhivotovskiy
Course Coordinator: Pedro Abdalla Teixeira

May 14, 2021

Try to solve the questions before looking to the answers. Every item must be proved rigorously. Starred problems are harder.

Problem 1

Consider i.i.d standard Gaussian random variable $g_1, \dots, g_n$ and a unit vector $a = (a_1, \dots, a_n)$.

(a) Use the rotation invariance property of the Gaussian random variable to derive a concentration inequality for the sum $\sum_{i=1}^{n} a_i g_i$.

(b) Use the Chernoff Method to derive a similar concentration inequality for the sum $\sum_{i=1}^{n} a_i g_i$. Compare with the inequality in letter "a".

Solution 1

(a) Observe that, by the invariance property of the Gaussian distribution, $\sum_{i=1}^{n} a_i g_i$ is Gaussian random variable g^* with mean zero and variance one. Therefore by a standard estimate for the tail of a Gaussian random variable we have

$$\mathbb{P}(|g^*| \geq t) \leq \min(1, \sqrt{\frac{2}{\pi}\frac{1}{t}})e^{-t^2/2}.$$

(b) By independence, we can write

$$\mathbb{E}e^{\lambda \sum_{i=1}^{n} a_i g_i} = \prod_{i=1}^{n} \mathbb{E}^{\lambda a_i g_i} = \prod_{i=1}^{n} e^{\lambda^2 a_i^2/2} = e^{\lambda^2/2}.$$

By the Chernoff method, we have

$$\mathbb{P}(\sum_{i=1}^{n} g_i a_i \geq t) \leq \inf_{\lambda \geq 0} e^{-\lambda t} e^{\lambda^2/2}.$$

For $\lambda = t$, we get $\mathbb{P}(\sum_{i=1}^{n} g_i a_i \geq t) \leq e^{-t^2/2}$. Finally by union bound, $\mathbb{P}(|\sum_{i=1}^{n} g_i a_i| \geq t) \leq 2e^{-t^2/2}$. We observe that both concentration inequalities have similar tail decay, the first inequality, provides a sharper estimate but it uses a crucial property exclusive for Gaussian distribution, while the second concentration inequality only requires a bound in the moment generating function, observe that such bound can be satisfied for other distributions (Check it!).

Problem 2

Show that the Sauer-Shelah lemma is tight, i.e, give an example of a family of events $\mathcal{A}$ with VC-dimension equal to d such that the shatter function is $\mathcal{S}_{\mathcal{A}}(n) = \sum_{i=1}^{d} \binom{n}{i}$.

Solution 2

Consider the family of binary sequences with at most d ones. Observe that for the growth function is $\sum_{j=1}^{d} \binom{n}{j}$, to see this observe that for every sequence of n "bits" (zeros and ones) with $i \leq d$ ones we have $\binom{n}{i}$ possibilities. For every $i \in [d]$ the sequences are different, so we have exactly $\sum_{i=0}^{d} \binom{n}{i}$. The shatter function $\mathcal{S}(d) = \sum_{i=0}^{d} \binom{d}{i} = 2^d$, so such family has VC-dimension at least d. Finally, $\mathcal{S}(d+1) = \sum_{i=0}^{d} \binom{d+1}{i} < 2^{d+1}$, therefore the VC-dimension is exactly d.

Problem 3

Compute the VC-dimension of the following classes

(a) Indicator function of sets of the form $[a, b] \cup [c, d]$ in $\mathbb{R}$.

(b) Indicator function of all circles in $\mathbb{R}^2$.

(c) All sets of the form $\{x \in \mathbb{R} : \sin(xt) \geq 0\}$ for all $t > 0$.

Solution 3

(a) We claim that the VC-dimension is four. Pick four points in the real line, namely $x_1 < x_2 < x_3 < x_4$. All the procedure will be with respect to this order. We assume $a < b < c < d$. We choose a to be equal to the first x_i with label 1 (if none exits just choose an $a > x_4$) and b to be strictly less than the first x_i with label 0 (if none exits choose an $b > x_4$), we repeat the process for c and d. It is easy to check that $\{x_1, \dots, x_4\}$ is shattered by such family of intervals. This proves that the VC-dimension of the class is at least 4. Now consider five points $x_1 < x_2 < x_3 < x_4 < x_5$ with alternating labels, i.e, $x_1 = 1$, $x_2 = 0$, $x_3 = 1$, $x_4 = 0$, $x_5 = 1$. To shatter x_1, x_2, x_3 we need two intervals, because no single interval can contain x_1 and x_3 without containing x_2. But the second interval cannot contain x_4 and we need another interval to contain x_5, i.e, we need at least three closed intervals to shatter $\{x_1, \dots, x_5\}$. Since this set of five points is arbitrary we can conclude that no set with five points can be shattered by the union of two closed intervals.

(b) It is easy to check that three points $\{x_1, x_2, x_3\}$ aligned in the x-axis (say), $x_1 = (-1, 0)$, $x_2 = (0, 0)$ and $x_3 = (1, 0)$ can be shattered by circles in $\mathbb{R}^2$. So the VC-dimension is at least three. Now we analyse four generic points $\{x_1, \dots, x_4\}$. Let K be the convex hull of such points, if one of the points do not lie in the boundary of K then the set cannot be shattered by the circles, indeed if the point in the interior of K has label 0 while the boundary points have label 1, then the circle must contain the interior point. Now we assume that all points x_i lie on the boundary of K. Without loss of generality, we assume that $x_1, \dots x_4$ are oriented in a clock-wise order and the distance between the opposite points x_1 and x_3 is smaller than the distance between the opposite points x_2 and x_4. If x_2 and x_4 have label one while x_1 and x_3 have label zero, there is no circle passing through x_2 and x_4 and not passing through x_1 and x_3. To see this, let H be a hyperplane induced by the line passing through x_1 and x_3. If the center of the circle O lies in the same region as x_2, then the circle contains either x_1 or x_3. The same holds if O lies in the region of x_4. In other words the circle must contain either x_1 or x_3 because the distance between x_2 and x_4 is larger than the distance between x_1 and x_3. We conclude that the VC-dimension is exactly three.

(c) We now show that the family $\{2^{-j} : j \in [m]\}$ can be shattered by the family of sines. It will imply that the VC-dimension of the class is infinite. We will work with labels ± 1 for simplicity. Consider $\{x_1, \ldots, x_m\}$ with labels $\{y_1, \ldots, y_m\}$. Now we choose $t := \pi(1 + \sum_{i=1}^m 2^i z_i)$ where $z_i = \frac{1-y_i}{2}$. Observe that

$$tx_j = t2^{-j} = \pi(2^{-j} + \sum_{i=1}^{m} 2^{i-j} z_i) = \pi(2^{-j} + \sum_{i=1}^{j-1} 2^{i-j} z_i + z_j + \sum_{i=1}^{m-j} 2^i z_i).$$

The latter is a multiple of 2π, so it can be dropped because of the periodicity of the sine function. The remaining term can be upper bounded by $\pi(\sum_{i=1}^{j} 2^{-i} + z_i) < \pi(1 + z_i)$. Similarly it can be (strictly) lower bounded by πz_i. So if $y_j = 1$, the sign of the sine is positive, otherwise the sign is negative. We correctly classify all points x_j, $j \in [m]$. Since m is arbitrary we can conclude that the class can shatter sets of arbitrary sizes. By definition it has infinite VC-dimension.