CS5014 Foundations of Data Science & Machine Learning

Quiz 02 — Solutions

March 31, 2021 — 8.30 - 9.30 AM

Instructions

1. This is a **zoom proctored** exam. Please adjust your seating so that **your face, hands, answer book and the mobile phone that you will use to scan the sheets are always in the webcam view**. Do not leave your seat or talk to anyone during the exam.
2. Write your answer on plain paper with your **name and roll number on the first sheet.**
3. This is a **closed book** exam. Do not refer to any books, notes, the Internet or any other person during the exam.
4. You can take **maximum 5 minutes after the exam to scan** the sheets into a **single PDF** file and upload to Moodle. Submissions made after 9:45 AM will be evaluated only if there is a genuine reason for the delay.

Questions

Question 1. Show that any finite hypothesis class is PAC learnable.

Definition 1 (PAC learnable)**.** A hypothesis class H over a domain X is called *PAC learnable* if, for every $\epsilon, \delta > 0$, there is an $n = n(H, \epsilon, \delta)$ such that for every binary function f on X and every probability distribution D on X,

$$P_{S \sim D^n}\left[\exists h \in H : (E_S(h, f) = 0) \wedge (E_D(h, f) > \epsilon)\right] \leq \delta. \tag{1}$$

Pick $\epsilon > 0$, $\delta > 0$, f and D arbitrarily and fix them. Let $H_\epsilon = \{h \in H :\ E_D(h, f) > \epsilon\}$. For any $h \in H_\epsilon$, the probability that a point from X sampled according to D will miss the region of disagreement between h and f equals $(1 - E_D(h, f)) \leq (1 - \epsilon)$. Since the n samples are chosen independently, the probability that this will happen for every sample is at most $(1-\epsilon)^n \leq e^{-\epsilon n}$. That is $P_{S \sim D^n}\left[E_S(h, f) = 0\right] \leq e^{-\epsilon n}$. Hence

$$\begin{aligned}
P_{S \sim D^n}\left[\exists h \in H : (E_S(h, f) = 0) \wedge (E_D(h, f) > \epsilon)\right] &\leq P_{S \sim D^n}\left[\exists h :\in H_\epsilon : E_S(h, f) = 0\right] \\
&\leq |H_\epsilon|\, P_{S \sim D^n}\left[E_S(h, f) = 0\right] \text{ (Union Bound)} \\
&\leq |H_\epsilon|\, e^{-\epsilon n} \\
&\leq |H|\, e^{-\epsilon n}.
\end{aligned}$$

Now one can either substitute an appropriate n (like $\frac{1}{\epsilon} \ln \frac{|H|}{\delta}$) and show that the above bound is at most δ for that n (and hence beyond). Otherwise, you can work backwards on an implication chain to get the appropriate n.

Question 2. Show that for all positive integers k, we have $e^n \geq n^k$ whenever $n \geq 2k \ln k$.

We will show that $n \geq k \ln n$ whenever $n \geq 2k \ln k$. Consider the function $f : (0, \infty) \to \mathbb{R}$ defined by $f(x) = x - k \ln x$. Then $f'(x) = 1 - k/x$ which is positive for all $x > k$. Hence $f(x)$ is increasing in (k, ∞). Hence it suffices for our purposes to show that $f(2k \ln k) \geq 0$.

$$\begin{aligned} f(2k \ln k) &= 2k \ln k - k \ln(2k \ln k) \\ &> 2k \ln k - k \ln(k^2) && (\forall x \in (0, \infty),\ \ln x \leq x/e < x/2) \\ &= 0. \end{aligned}$$

One can prove that $\forall x \in (0, \infty),\ \ln x \leq x/e$ by drawing their graphs. The line $y = x/e$ will be the tangent to the curve $y = \ln x$ at $(e, 1)$.

Question 3. Show that the VC-dimension of triangles in $\mathbb{R}^2$ is 7. You can use the facts below without proving them.

Fact 1. If a set of points P is shattered by triangles, then no point $p \in P$ is contained in the convex hull of $P \setminus \{p\}$.

Fact 2. A triangle and a convex polygon cross in at most 6 points.

In order to show that VC-dim of triangles is ≥ 7, it is enough to show one set S of 7 points in $\mathbb{R}^2$ which can be shattered by triangles. We will choose S to be 7 distinct points on the unit circle. We have to show that for each of the 2^7 possible binary labellings of S, there is a triangle which contains all the points labelled $+1$ and none of the points labelled -1. The usual proof is to group the cases based on the number of $+1$-labelled points in S. You still have to analyse two or three configurations among cases where the number $+1$-labelled points are the same. For each configuration you can give an easy picture proof. The common error is when you miss out some cases. A more efficient proof (from here) follows. In any binary labelling of S, there are at most 3 consecutive stretches of -1-labelled points. This is obvious if S has ≤ 3 $+1$ points. When S has more than three 1 points, then the number of separators (-1 points) is at most 3 and hence number of consecutive stretches of $+1$ points is at most 3. Each such stretch can be separated by a line and these lines will intersect only outside of the unit circle. The triangle formed by these three lines will separate the $+1$ points from the -1 points.

In order to show that VC-dim of triangles is < 8 we have to argue that no set S of 8 points can be shattered by triangles. If the S is not convex, then the labelling which assigns $+1$ to all points of S which lie on the convex hull of S and -1 to the interior ones cannot be separated by a triangle. Since we have handled this case, we can assume that all the 8 points of S define a convex polygon P. Now label the 8 points $+1$ and -1 alternatively. Any triangle separating this configuration should intersect P eight times, which contradicts Fact 2.

Total points: $3 \times 10 = 30$.