36-702 Homework 1 Solution

Thanks to William Bishop and Rafael Stern for providing their solutions.

Problem 1

(a) Let $n(j) = \sum_i I_{\{j\}}(x_i)$,

$$L(\theta) = \theta^{n(1)} \cdot \left(\frac{\theta}{2}\right)^{n(2)} \cdot \left(\frac{\theta}{3}\right)^{n(3)} \cdot \left(\frac{6-11\theta}{6}\right)^{n(4)} \propto$$

$$\propto \theta^{n(1)+n(2)+n(3)} \cdot (6-11\theta)^{n(4)}$$

Thus, there exists a constant k such that:

$$l(\theta) = k + (n - n(4))\log(\theta) + n(4)\log(6-11\theta)$$

$$\frac{dl(\theta)}{d\theta} = \frac{n-n(4)}{\theta} - \frac{11n(4)}{6-11\theta} = \frac{(n-n(4))(6-11\theta) - 11n(4)\theta}{\theta(6-11\theta)}$$

Hence, $\frac{dl(\theta)}{d\theta} = 0$ if and only if:

$$6(n - n(4) = 11n\theta$$

$$\theta = \frac{6(n-n(4))}{11n}$$

Since $L(0) = L(6/11) = 0$ and $\Theta \cup \{0, 6/11\}$ is a closed set, by Weierstrass's Theorem, $\frac{6(n-n(4))}{11n}$ maximizes L.

(b) From the previous item:

$$\frac{dl(\theta)}{d\theta} = \frac{n - n(4)}{\theta} - \frac{11n(4)}{6 - 11\theta}$$

Thus,

$$\frac{d^2l(\theta)}{d\theta^2} = -\frac{n - n(4)}{\theta^2} - \frac{121n(4)}{(6 - 11\theta)^2} =$$

Finally, $I(\theta) = -E(\frac{d^2l(\theta)}{d\theta^2}) =$

$$= \frac{n - \frac{n(6-11\theta)}{6}}{\theta^2} + \frac{121\frac{n(6-11\theta)}{6}}{(6 - 11\theta)^2} =$$

$$= \frac{11n}{6\theta} + \frac{121n}{6(6 - 11\theta)} = \frac{11n(6 - 11\theta) + 121n\theta}{6\theta(6 - 11\theta)} = \frac{11n}{\theta(6 - 11\theta)}$$

(c) We know that $\frac{\hat{\theta}-\theta}{\sqrt{I(\hat{\theta})}} \xrightarrow{\mathcal{L}} N(0,1)$. Hence an asymptotic $1 - \alpha$ confidence interval for θ is:

$$\left[\hat{\theta} - z_{1-\alpha/2}\sqrt{I(\hat{\theta})}, \hat{\theta} + z_{1-\alpha/2}\sqrt{I(\hat{\theta})}\right]$$

That is,

$$\left[\frac{6}{11}\left(1 - \frac{n(4)}{n}\right) - z_{1-\alpha/2}\frac{11\sqrt{n}}{6\sqrt{\frac{n(4)}{n}\left(1 - \frac{n(4)}{n}\right)}}, \frac{6}{11}\left(1 - \frac{n(4)}{n}\right) + z_{1-\alpha/2}\frac{11\sqrt{n}}{6\sqrt{\frac{n(4)}{n}\left(1 - \frac{n(4)}{n}\right)}}\right]$$

(d) From item (a), $\hat{\theta} = \frac{6}{11}\left(1 - \frac{n(4)}{n}\right)$. By the LGN, $\frac{n(4)}{n} \xrightarrow{P} \frac{6-11\theta}{6}$. Thus, by the continuous mapping theorem:

$$\hat{\theta} \xrightarrow{P} \frac{6}{11}\left(1 - \frac{6 - 11\theta}{6}\right) = \theta$$

Problem 2

For any $t \in \mathbb{R}$ define $f_t(z) = \text{sign}(\sin(tz))$. Let $\mathcal{F} = \{f_t : t \in \mathbb{R}\}$. Show that $\mathcal{F}$ has infinite VC dimension. Hint: consider a set of points like $\{1/2, 1/4, \ldots, 1/2^n\}$.

Proof. Define $z_i = \frac{1}{2^i}$ and define $\mathcal{Z}_n = \{z_i : i \in [1, 2, \ldots n]\}$. Note that if $f_t(z_i) = 1$ it must be that $\text{sign}(\sin(tz_i)) = 1$, which implies that:

$$\begin{aligned} \sin(tz_i) &> 0 \\ tz_i &> sin^{-1}(0) \end{aligned}$$

This implies that

$$0 < tz_i < \pi \text{ or } 2\pi < tz_i < 3\pi \text{ or } 4\pi < tz_i < 5\pi \ldots$$

Written succinctly, it must be that $0 + a_i 2\pi < tz_i < \pi + a_i 2\pi, a_i \in [0, 1, 2, \ldots, \infty)$.
Now, consider the case when $f_t z_i = -1$. In this case it must be that:

$$\begin{aligned} sin(tz_i) &< 0 \\ tz_i &< sin^{-1}(0) \end{aligned}$$

This implies that:

$$\pi < tz_i < 2\pi \text{ or } 3\pi < tz_i < 4\pi \text{ or } 5\pi < tz_i < 6\pi \ldots$$

Written succinctly, it must be that $\pi + a_i 2\pi < tz_i < 2\pi + a_i 2\pi, a_i \in [0, 1, 2, \ldots, \infty)$.
Now, let $y_i = f_t(z_i)$ and let $\mathcal{Y}_n = \{y_i : i \in [1, \ldots n]\}$. For a given $\mathcal{Y}_n$ we can write a set of n inequalities:

$$\begin{cases} 0 + a_i 2\pi < tz_i < \pi + 2a_i\pi, & y_i = 1 \\ \pi + a_i 2\pi < tz_i < 2\pi + 2a_i\pi, & y_i = -1 \end{cases}$$

Substituting for z_i, this becomes:

$$\begin{cases} 0 + a_i 2\pi < t\frac{1}{2^i} < \pi + 2a_i\pi, & y_i = 1 \\ \pi + a_i 2\pi < t\frac{1}{2^i} < 2\pi + 2a_i\pi, & y_i = -1 \end{cases}$$

Let $t = k\pi$, we can then write the set of equations as:

$$\begin{cases} 0 + a_i 2\pi < k\pi\frac{1}{2^i} < \pi + 2a_i\pi, & y_i = 1 \\ \pi + a_i 2\pi < k\pi\frac{1}{2^i} < 2\pi + 2a_i\pi, & y_i = -1 \end{cases}$$

and this is equal to:

$$\begin{cases} 0 + a_i 2 < k\frac{1}{2^i} < 1 + 2a_i, & y_i = 1 \\ 1 + a_i 2 < k\frac{1}{2^i} < 2 + 2a_i, & y_i = -1 \end{cases}$$

We add 1 to the top equation to get:

$$\begin{cases} 1 + a_i 2 < 1 + \frac{k}{2^i} < 2 + 2a_i, & y_i = 1 \\ 1 + a_i 2 < \frac{k}{2^i} < 2 + 2a_i, & y_i = -1 \end{cases}$$

Finally, for convenience, subtract $2a_i$ from both equations to get:

$$\begin{cases} 1 < 1 + \frac{k}{2^i} - 2a_i < 2, & y_i = 1 \\ 1 < \frac{k}{2^i} - 2a_i < 2, & y_i = -1 \end{cases}$$

Now, let $\mathcal{Y} = \{y_1, \ldots, y_n\}$ be a set of any arbitrary labelings for the points in $\mathcal{Z}_n$. Note, that there are 2^n such labelings. I will now prove that it is possible to find a solution to the set of equations defined directly above for all such labelings.

In general, if $y_i = 1$, we can say the following:

$$1 < 1 + \frac{k}{2^i} - 2a_1 < 2$$
$$2^i < 2^i + k - 2^{i+1}a_1 < 2^{i+1}$$
$$0 < k - a_1 2^{i+1} < 2^{i+1} - 2^i$$

And if $y_i = -2$, we can say:

$$1 < \frac{k}{2^i} - 2a_1 < 2$$
$$2^i < k - 2^{i+1}a_1 < 2^{i+1}$$

This is a very important pattern. For a fixed i, we have an alternating set of ranges for which k can be in if $y_i = 1$, ie: $(0, 2^{i+1} - 2^i), (1 \times 2^{i+1}, 1 \times 2^{i+1} + 2^{i+1} - 2^i), (2 \times 2^{i+1}, 2 \times 2^{i+1} + 2^{i+1} - 2^i), \ldots$. We also have an alternating set of ranges for which k can be in if $y_i = 1$. Notice that these ranges are situated between those for $y_i = 1$, ie: $(2^i, 2^{i+1}), (1 \times 2^{i+1} + 2^i, 1 \times 2^{i+1} + 2^{i+1}), (2 \times 2^{i+1} + 2^i, 2 \times 2^{i+1} + 2^{i+1}), \ldots$.

Further, notice what happens as we move from ranges for i and ranges for $i+1$. The ranges for $i+1$ are twice as big as the ranges for i. Additionally, the ranges for k when $y_{i+1} = 1$ will cover the ranges for k when $y_i = 1$ *and* $y_i = -1$. This also holds for the ranges when $y_{i+1} = -1$. Thus, no matter what the label for y_i, it will always be possible to find a value of k that correctly predicts the label for y_i and y_{i+1}. This relationship holds for all i, so it must be that for any arbitrary set $\mathcal{Y}_n$, we can find a $k = t\pi$, such that f_t correctly maps all $z \in \mathcal{Z}_n$ to $y \in \mathcal{Y}_n$. This is equivalent to saying $s(\mathcal{F}, n) = 2^n$ for all n, which implies the VC dimension for $\mathcal{F}$ is infinity.

□

Problem 3

(a) We wish to find an ϵ such that $P(|\bar{X} - \mu| > \epsilon) \leq \delta$. By Bernstein's Inequality, it is sufficient to take an ϵ such that:

$$2exp\left(-\frac{n\epsilon^2}{2\sigma^2 + 2c\epsilon/3}\right) \leq \delta$$

$$\frac{n\epsilon^2}{2\sigma^2 + 2c\epsilon/3} \geq \log(2/\delta)$$

$$\epsilon^2 - \frac{2c\log(2/\delta)\epsilon}{3n} - \frac{2\sigma^2\log(2/\delta)}{n} \geq 0$$

Hence, it is sufficient to take ϵ such that:

$$\epsilon \geq \frac{c\log(2/\delta)}{3n} + \sqrt{\frac{\left(\frac{2c\log(2/\delta)}{3n}\right)^2 + \frac{8\sigma^2\log(2/\delta)}{n}}{4}}$$

Remembering that $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$, it is sufficient to take ϵ such that:

$$\epsilon \geq \frac{2c\log(2/\delta)}{3n} + \sqrt{\frac{2\sigma^2\log(2/\delta)}{n}}$$

(b) We wish to show that:

$$\forall \epsilon > 0, \exists C > 0 : P(n|\bar{X}_n - \mu_n| > C) \leq \epsilon$$

First, observe that since X_i are dicotomic, $|X_i| \leq 1$. Next, if we call $P(Y_i \in A_n) = p_n$, $\sigma^2 \leq p_n(1-p_n) \leq p_n$. Also call $M = \sup f$ and observe that, $p_n \leq \frac{M}{n}$ and, thus, $\sigma^2 \leq \frac{M}{n}$. Hence, by Bernstein's Inequality:

$$P(|\bar{X}_n - \mu_n| > \frac{C}{n}) \leq 2exp\left(-\frac{\frac{C^2}{n}}{\frac{2M}{n} + \frac{2C}{3n}}\right)$$

$$P(|\bar{X}_n - \mu_n| > \frac{C}{n}) \leq 2exp\left(-\frac{C}{2M/C + 2/3}\right)$$

The result follows observing that $2exp\left(-\frac{C}{2M/C+2/3}\right) \xrightarrow{C \mapsto \infty} 0$.

Problem 4

(Rademacher Complexity). Let $\mathcal{F} = \{f_1, \ldots, f_N\}$ where each f is a binary function: $f(x) \in \{0,1\}$. Show that

$$\mathcal{R}_n(\mathcal{F}) \leq 2\sqrt{\frac{\log N}{n}}$$

Proof. From theorem 42.33 in the notes, we know that:

$$\mathcal{R}_n(\mathcal{F}) \leq \sqrt{\frac{2\log(s(\mathcal{F},n))}{n}}$$

For a finite set of functions $\mathcal{F} = \{f_1, \ldots, f_N\}$ it must be that $s(\mathcal{F}, n) \leq N$ as it is impossible to produce more unique labelings than there are functions in the set.

Thus, it immediately follows that:

$$\mathcal{R}_n(\mathcal{F}) \leq \sqrt{\frac{2\log(N)}{n}} \leq 2\sqrt{\frac{\log N}{n}}$$

□