Dongrong Li
STAT 6050
February 19, 2022

HOMEWORK 1

1. Problem 1

1.1. **Problem 1.1.** Let Z be $Z_1, ..., Z_n$. σ, ρ are also defined respectively.
To show the left hand side,

$$\begin{aligned} E\sup_{h\in\mathcal{H}}|Gauss_n(h)| &= E\sup_{h\in\mathcal{H}}|\frac{1}{n}\sum_{i=1}^{n}\sigma_i h(Z_i)| \\ &= E_{\sigma,\rho,Z}\sup_{h\in\mathcal{H}}|\frac{1}{n}\sum_{i=1}^{n}|\sigma_i|\rho_i h(Z_i)| \\ &\geq E_{\rho,Z}\sup_{h\in\mathcal{H}}E_\sigma|\frac{1}{n}\sum_{i=1}^{n}|\sigma_i|\rho_i h(Z_i)| \quad (1) \\ &\geq E_{\rho,Z}\sup_{h\in\mathcal{H}}|\frac{1}{n}\sum_{i=1}^{n}E_\sigma|\sigma_i|\rho_i h(Z_i)| \quad (2) \\ &= E_{\rho,Z}E_\sigma|\sigma_i|\sup_{h\in\mathcal{H}}|\frac{1}{n}\sum_{i=1}^{n}\rho_i h(Z_i)| \\ &= c_0 E_{\rho,Z}\sup_{h\in\mathcal{H}}|Rad_n(h)| \end{aligned}$$

Here $c_0 = E|\sigma|$ for $\sigma \sim N(0,1)$. (1) and (2) come from Jensen's inequality.

To show the right hand side, first notice that for any $n \geq 1$, $\log n \geq d\sqrt{\log n}$, where $d = \frac{1}{\sqrt{\log 2}}$.

$$\begin{aligned} E\sup_{h\in\mathcal{H}}|Gauss_n(h)| &= E\sup_{h\in\mathcal{H}}|\frac{1}{n}\sum_{i=1}^{n}\sigma_i h(Z_i)| \\ &= E_{\sigma,\rho,Z}\sup_{h\in\mathcal{H}}|\frac{1}{n}\sum_{i=1}^{n}|\sigma_i|\rho_i h(Z_i)| \\ &= E_\sigma E_{\rho,Z}\sup_{h\in\mathcal{H}}|\frac{1}{n}\sum_{i=1}^{n}|\sigma_i|\rho_i h(Z_i)| \\ &= E_\sigma E_{\rho,Z}\sup_{h\in\mathcal{H}}|Rad_n(\sigma h)| \\ &\leq E_\sigma||\sigma||_\infty E_{\rho,Z}\sup_{h\in\mathcal{H}}|Rad_n(h)| \quad (1) \\ &\leq \sqrt{2\log n}E_{\rho,Z}\sup_{h\in\mathcal{H}}|Rad_n(h)| \\ &= 2d(\log n)E\sup_{h\in\mathcal{H}}|Rad_n(h)| \\ &= c_1(\log n)E\sup_{h\in\mathcal{H}}|Rad_n(h)| \end{aligned}$$

Where $c_1 = 2d$. (1) comes from the contraction inequality in Lecture 4.

1.2. **Problem 1.2.** The result from problem 1.1 suggests that the Rademacher complexity is upper bounded by Gaussian complexity and the Gaussian complexity can further be bounded by Rademacher complexity up to $\log n$. Asymptotically, $\frac{E\sup_{h\in\mathcal{H}|Gauss_n(h)|}}{E\sup_{h\in\mathcal{H}|Rad_n(h)|}} = o(n)$

2. Problem 2

2.1. **Problem 2.1.** The Bousquet bound suggests that:

$$P(||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}} - E||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}} \geq \sqrt{\frac{2\log(1/\delta)}{n}(\sigma^2_{\mathcal{H}} + E||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}})} + \frac{U\log(1/\delta)}{3n}) \leq \delta$$

Now, for any $\epsilon_n > 0$, let

$$\delta = \exp{(-\frac{n\epsilon_n^2}{2(\sigma^2_{\mathcal{H}} + E||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}} + U\epsilon_n/3)})}$$

By taking logarithm, we obtain:

$$n\epsilon_n^2 = 2\log(1/\delta)(\sigma^2_{\mathcal{H}} + E||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}} + U\epsilon_n/3)$$

Solving this quadratic form, we obtain

$$\epsilon_n = \frac{U}{3n}\log(1/\delta) \pm \sqrt{\frac{U^2(\log(1/\delta))^2}{9n^2} + \frac{2}{n}\log(1/\delta)(\sigma^2_{\mathcal{H}} + E||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}})}$$

Because $\epsilon_n > 0$, the solution of ϵ_n should be:

$$\begin{aligned}\epsilon_n &= \frac{U}{3n}\log(1/\delta) + \sqrt{\frac{U^2(\log(1/\delta))^2}{9n^2} + \frac{2}{n}\log(1/\delta)(\sigma^2_{\mathcal{H}} + E||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}})} \\ &\geq \frac{U}{3n}\log(1/\delta) + \sqrt{\frac{2}{n}\log(1/\delta)(\sigma^2_{\mathcal{H}} + E||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}})}\end{aligned}$$

Therefore, we obtain:

$$\begin{aligned}&P(||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}} - E||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}} \geq \epsilon_n) \\ &\leq P(||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}} - E||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}} \geq \frac{U}{3n}\log(1/\delta) + \sqrt{\frac{2}{n}\log(1/\delta)(\sigma^2_{\mathcal{H}} + E||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}})}) \\ &\leq \delta = \exp{(-\frac{n\epsilon_n^2}{2(\sigma^2_{\mathcal{H}} + E||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}} + U\epsilon_n/3)})}\end{aligned}$$

Where the second inequality is from Bousquet bound.

Furthermore, given that $\epsilon_n \geq 2E||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}}$, we have

$$\begin{aligned}P(||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}} \geq \epsilon_n) &= P(||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}} - E||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}} \geq \epsilon_n - E||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}}) \\ &\leq P(||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}} - E||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}} \geq \frac{\epsilon_n}{2}) \\ &\leq \exp{(-\frac{n(\epsilon_n/2)^2}{2(\sigma^2_{\mathcal{H}} + E||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}} + U\epsilon_n/3)})} \\ &= \exp{(-\frac{n(\epsilon_n)^2}{8(\sigma^2_{\mathcal{H}} + E||\mathbb{P}-\mathbb{P}_n||_{\mathcal{H}} + U\epsilon_n/3)})} \\ &\leq \exp{(-\frac{n(\epsilon_n)^2}{8(\sigma^2_{\mathcal{H}} + \epsilon_n/2 + U\epsilon_n/3)})} \\ &= \exp{(-\frac{n(\epsilon_n)^2}{8(\sigma^2_{\mathcal{H}} + (1/2 + U/3)\epsilon_n)}}\end{aligned}$$

3. Problem 3

3.1. Problem 3.1.

If $t \geq 1$, let

$$P(X = t^2) = P(X = -t^2) = \frac{1}{2t^2}$$

and

$$P(X = 0) = 1 - 1/t^2$$

In this case, $EX = 0$ and $Var(X) = t^2$, so

$$P(|X - EX| \geq t) = P(|X| \geq t) = P(|X| \geq 1) = 1 = Var(X)/t^2$$

For the case $0 < t < 1$, let

$$P(X = t) = P(X = -t) = 1/2$$

We can easily check that $EX = 0$ and $Var(X) = t^2$, In this case,

$$P(|X - EX| \geq t) = P(|X| \geq t) = 1 = Var(X)/t^2$$

3.2. Problem 3.2.

Let $\tilde{Z}_i$ be a generic independent copy of Z

$$\begin{aligned}
LHS &= \frac{1}{2} E \sup_{h \in \mathcal{H}} |\frac{1}{n} \sum_{i=1}^{n} \rho_i (h(Z_i) - Eh(Z_i))| \\
&= \frac{1}{2} E \sup_{h \in \mathcal{H}} |\frac{1}{n} \sum_{i=1}^{n} \rho_i (h(Z_i) - \tilde{E}h(\tilde{Z}_i))| \\
&\leq \frac{1}{2} E \sup_{h \in \mathcal{H}} \tilde{E} |\frac{1}{n} \sum_{i=1}^{n} \rho_i (h(Z_i) - h(\tilde{Z}_i))| \quad (1) \\
&\leq \frac{1}{2} E\tilde{E} \sup_{h \in \mathcal{H}} |\frac{1}{n} \sum_{i=1}^{n} \rho_i (h(Z_i) - h(\tilde{Z}_i))| \quad (2) \\
&= \frac{1}{2} E\tilde{E} \sup_{h \in \mathcal{H}} |\frac{1}{n} \sum_{i=1}^{n} (h(Z_i) - h(\tilde{Z}_i))| \quad (3) \\
&= \frac{1}{2} E\tilde{E} \sup_{h \in \mathcal{H}} |\frac{1}{n} \sum_{i=1}^{n} (h(Z_i) - Eh(Z_i) + Eh(Z_i) - h(\tilde{Z}_i))| \\
&\leq \frac{1}{2} E_{Z,\tilde{Z},\rho} \sup_{h \in \mathcal{H}} |h(Z_i) - Eh(Z_i)| + |Eh(Z_i) - h(\tilde{Z}_i)| \\
&= E_{Z,\tilde{Z},\rho} \sup_{h \in \mathcal{H}} |h(Z_i) - Eh(Z_i)| \\
&= RHS
\end{aligned}$$

Where (1) and (2) comes from Jensen's inequality and (3) comes from symmetrization.

3.3. Problem 3.3.

Let $Z \in R^n$ be $Z = (Z_1, ..., Z_n)$

$$
\begin{aligned}
E\sup_{h\in\mathcal{H}}|\frac{1}{n}\sum_{i=1}^{n}\rho_i h(Z_i)| &= E\sup_{h\in\mathcal{H}}|\frac{1}{n}\sum_{i=1}^{n}\rho_i\beta^T Z_i| \\
&\leq E\sup_{h\in\mathcal{H}}\frac{1}{n}||\beta^T||_2||Z||_2 \quad (1) \\
&= E\frac{1}{n}||Z||_2\sup_{h\in\mathcal{H}}||\beta^T||_2 \\
&= E\frac{1}{n}||Z||_2
\end{aligned}
$$

Notice that (1) can be achieved by letting β at the same direction with Z.

Therefore, we conclude that the Rademacher complexity of $\mathcal{H}$ is $\frac{1}{n}E||Z||_2$.

This suggests that if the expectation of $||Z||_2$ grows linearly, then Rademacher complexity will be uniformly bounded. If the expecation of $||Z||_2$ grows at the rate of $\log n$, then Rademacher complexity goes to 0 asymptotically.

References