Homework 2

October 25, 2022

Name: Andor Vári-Kakas
Student number: 22-957-419
Student ID: avarikakas
I discussed the problems on a very high level with Kristóf Szabó (22-956-445).

1. (a)

$$\begin{aligned}
G_n(F_{B,s}(x_1^n)) &= \mathbb{E}_\omega \sup_f \frac{1}{n}\sum_{i=1}^n \omega_i f(x_i) \\
&= \mathbb{E}_\omega \sup_\theta \frac{1}{n}\sum_{i=1}^n \omega_i \langle \theta, x_i\rangle \\
&= \mathbb{E}_\omega \sup_\theta \frac{1}{n}\langle \theta, \sum_{i=1}^n \omega_i x_i\rangle \\
&= \frac{1}{\sqrt{n}}\mathbb{E}_\omega \sup_\theta \langle \theta, \frac{1}{\sqrt{n}}\sum_{i=1}^n \omega_i x_i\rangle \\
&= \frac{1}{\sqrt{n}}\mathbb{E}_\omega \sup_\theta \langle \theta, \frac{X^T\omega}{\sqrt{n}}\rangle \\
&= \frac{1}{\sqrt{n}}\mathbb{E}_\omega \max_{|S|=s} \sup_{\theta'\in\mathbb{R}^s:||\theta'||_2\le B} \langle \theta', \frac{X_S^T\omega}{\sqrt{n}}\rangle \text{ using the bound on the l0 norm of } \theta \\
&= \mathbb{E}_\omega \max_S \sup_{\theta'} ||\theta'||_2 ||\frac{X_S^T\omega}{n}||_2 \text{ using the Cauchy-Schwartz inequality} \\
&= B\mathbb{E}_\omega \max_S ||\frac{X_S^T\omega}{n}||_2 \text{ using the bound on the l2 norm of } \theta(')
\end{aligned}$$

(b) First let's bound the expectation of $||\omega_S||_2$.

$$\begin{aligned}
\mathbb{E}[||\omega_S||_2] &= \frac{1}{\sqrt{n}}\mathbb{E}[||X_S^T\omega||_2] \\
&= \mathbb{E}[\sqrt{\omega^T \frac{X_S X_S^T}{n}\omega}] \\
&= \mathbb{E}[\sqrt{\omega^T Q\Lambda Q^T\omega}] \text{ using the eigendecomposition of the real symmetric matrix } \frac{X_S X_S^T}{\sqrt{n}} \\
&\text{where } \Lambda \text{ is a diagonal matrix with the eigenvalues of } \frac{X_S X_S^T}{\sqrt{n}} \\
&= \mathbb{E}_\omega[\sqrt{(Q^T\omega)^T\Lambda(Q^T\omega}] \\
&= \mathbb{E}_{\omega'}[\sqrt{\omega'\Lambda\omega'^T}] \text{ as } Q^T\omega \text{ has the same distribution as } \omega \text{ since } Q \text{ is orthonormal}
\end{aligned}$$

Since X_S has rank $\le s$, Λ has at most s non-zero eigenvalues. Letting Λ_S correspond to the submatrix induced by non-zero eigenvalues (possibly filled up to size sxs), we get

$$
\begin{aligned}
&\leq \mathbb{E}_{\omega \in N(0,I^s)}[\sqrt{\omega \Lambda_S \omega^T}] \\
&\leq \mathbb{E}[\sqrt{C^2 ||\omega||_2^2}] \\
&= C\mathbb{E}\sqrt{||\omega||_2^2} \\
&\leq C\sqrt{\mathbb{E}||\omega||_2^2} \text{ by Jensen} \\
&\leq C\sqrt{\mathbb{E}\sum_{i=1}^{s}\omega_i^2} \\
&= C\sqrt{s} \text{ as } \sum_{i=1}^{s}\omega_i^2 \text{ is Chi-squared}
\end{aligned}
$$

Now I will show that $||\omega_S||$ is C-lipschitz.

$$
\begin{aligned}
|||\omega_S||_2 - ||\omega_S'||_2| &= \frac{|||X_S^T\omega||_2 - ||X_S^T\omega'||_2|}{\sqrt{n}} \\
&\leq \frac{|||X_S^T\omega - X_S^T\omega'||_2|}{\sqrt{n}} \text{ by triangle inequality} \\
&= \frac{||X_S^T\omega - X_S^T\omega'||_2}{\sqrt{n}} \\
&= \frac{||X_S^T(\omega - \omega')||_2}{\sqrt{n}} \\
&\leq C||\omega - \omega'||_2 \text{ using the same logic as in the second part of part a)}
\end{aligned}
$$

Now we will refer to a Gaussian concentration theorem (see e.g. Theorem 6.2 of https://people.math.ethz.ch/~abandeira/BandeiraSingerStrohmer-MDS-draft.pdf) to get the desired concentration bound, exploting that $||\omega_S||_2$ is C-lipschitz which is a prerequisite of the theorem:

$$
\begin{aligned}
P(||\omega_S||_2 \geq \sqrt{s}C + \delta) &\leq P(||\omega_S||_2 \geq \mathbb{E}[||\omega_S||_2] + \delta) \text{ using the bound on } \mathbb{E}[||\omega_S||_2] \text{ from earlier in this part} \\
&= P(||\omega_S||_2 - \mathbb{E}[||\omega_S||_2] \geq \delta) \\
&\leq e^{-\frac{\lambda^2}{2C^2}} \text{ using the cited theorem and that } ||\omega_S||_2 \text{ is } C\text{-lipschitz}
\end{aligned}
$$

Note that it also follows that $||\omega_S||_2$ is C-subgaussian.

(c)

$$\begin{aligned}
G_n(F_{B,s}(x_1^n)) &= B\mathbb{E}_\omega \max_S ||\frac{X_S^T\omega}{n}||_2 \text{ using part a)} \\
&= \frac{B}{\sqrt{n}}\mathbb{E}_\omega \max_S ||\frac{X_S^T\omega}{\sqrt{n}}||_2 \\
&= \frac{B}{\sqrt{n}}\mathbb{E}_\omega \max_S ||\omega_S||_2 \\
&= \frac{B}{\sqrt{n}}\mathbb{E}_\omega \max_S((||\omega_S||_2 - \mathbb{E}||\omega_S||_2) + \mathbb{E}||\omega_S||_2) \\
&\leq (\frac{B}{\sqrt{n}}\mathbb{E}_\omega \max_S((||\omega_S||_2 - \mathbb{E}||\omega_S||_2)) + C\sqrt{s} \text{ using the bound on the expection from earlier} \\
&\leq \frac{B}{\sqrt{n}}2C\sqrt{\log\binom{d}{s}} + C\sqrt{s} \text{ using Homework 1/3a)} \\
&\text{and that } ||\omega_S||_2 - \mathbb{E}[||\omega_S||_2] \text{ is } C\text{-subgaussian with 0 mean} \\
&\leq \frac{B}{\sqrt{n}}2C\sqrt{\log\left(\frac{de}{s}\right)^s} + C\sqrt{s} \text{ using a simple bound on the binomial coefficient} \\
&= 2BC\sqrt{\frac{s\log\left(\frac{de}{s}\right)}{n}} + C\sqrt{s} \\
&= O(BC\sqrt{\frac{s\log\left(\frac{de}{s}\right)}{n}})
\end{aligned}$$

(d)

$$\begin{aligned}
G_n(\tilde{F}_{B,s}(x_1^n)) &= \frac{1}{\sqrt{n}}\mathbb{E}_\omega \sup_\theta\langle\theta, \frac{X^T\omega}{\sqrt{n}}\rangle \text{ using part a)} \\
&= \frac{1}{\sqrt{n}}\mathbb{E}_\omega \sup_\theta\langle\frac{X\theta}{\sqrt{n}}, \omega\rangle \\
&= \frac{1}{\sqrt{n}}\mathbb{E}_\omega \max_{|S|=s}\sup_{\theta_S}\langle\frac{X\theta_S}{\sqrt{n}}, \omega\rangle \text{ where } \theta_S \text{ can only have non-zero values in the positions } S \\
&= \frac{1}{\sqrt{n}}\mathbb{E}_\omega \max_{|S|=s}\sup_{\theta_S}\langle\frac{X\theta_S}{\sqrt{n}}, P\omega\rangle \text{ where } P_S \text{ is a projection matrix onto the vector space of } X_S \\
&\text{because the remaining orthogonal components of } \omega \text{ result in a 0 dot-product due to the 0s in } \theta_S \\
&\leq \frac{1}{\sqrt{n}}\mathbb{E}_\omega \max_{|S|=s}\sup_{\theta_S}||\frac{X\theta_S}{\sqrt{n}}||_2||P_S\omega||_2 \text{ by Cauchy-Schwartz} \\
&\leq \frac{1}{\sqrt{n}}\mathbb{E}_\omega \max_{|S|=s}\sup_{\theta_S}B||P_S\omega||_2 \text{ by the constraint on } ||\frac{X\theta_S}{\sqrt{n}}||_2 \\
&= \frac{B}{\sqrt{n}}\mathbb{E}_\omega \max_{|S|=s}||P_S\omega||_2
\end{aligned}$$

Now note that since X_S has rank $\leq s$, P_S projects onto a subspace of dimension $\leq s$. Hence, $\mathbb{E}[||P_S\omega||_2] = \sqrt{s}$. Also, $||P_S\omega||_2$ can be easily shown to be 1-lipschitz (*a factor C less than before!*). Combining these and using an analogous argument to part c), we get the desired result:

$$G_n(\tilde{F}_{B,s}(x_1^n)) = O(B\sqrt{\frac{s\log\left(\frac{de}{s}\right)}{n}})$$

2. (a)

$$\begin{aligned}
G_n(T) &= \frac{1}{n}\mathbb{E}_\omega \sup_\theta \sum_{i=1}^n \omega_i\theta_i \\
&\geq \frac{1}{n}\mathbb{E}_\omega \sum_{i=1}^n \omega_i(\arg\max_\theta\langle \mathrm{sign}(\omega), \theta\rangle)_i \text{ by picking a single } \theta \text{ instead of sup} \\
&= \frac{1}{n2^n}\mathbb{E}_\omega[\sum_\epsilon \sum_{i=1}^n \omega_i(\arg\max_\theta\langle \epsilon, \theta\rangle)_i \mid \mathrm{sign}(\omega) = \epsilon] \text{ by conditional expectation and symmetry of } \omega \\
&= \frac{1}{n2^n}\sum_\epsilon \sum_{i=1}^n \mathbb{E}_\omega[\omega_i \mid \mathrm{sign}(\omega) = \epsilon](\arg\max_\theta\langle \epsilon, \theta\rangle)_i \\
&= \frac{1}{n2^n}\sum_\epsilon \sum_{i=1}^n \mathbb{E}_{\omega_i}[\omega_i \mid \mathrm{sign}(\omega) = \epsilon](\arg\max_\theta\langle \epsilon, \theta\rangle)_i \\
&= \frac{1}{n2^n}\sum_\epsilon \sum_{i=1}^n \mathbb{E}_{\omega_i}[|\omega_i|]\epsilon_i(\arg\max_\theta\langle \epsilon, \theta\rangle)_i \\
&= \frac{1}{n2^n}\sum_\epsilon \sum_{i=1}^n \sqrt{\frac{2}{\pi}}\epsilon_i(\arg\max_\theta\langle \epsilon, \theta\rangle)_i \text{ using the expectation of the half-normal distribution} \\
&= \sqrt{\frac{2}{\pi}}\frac{1}{n2^n}\sum_\epsilon \max_\theta\langle \epsilon, \theta\rangle \\
&= \sqrt{\frac{2}{\pi}}\frac{1}{n}\sum_\epsilon \frac{1}{2^n}\max_\theta\langle \epsilon, \theta\rangle \\
&= \sqrt{\frac{2}{\pi}}\frac{1}{n}\mathbb{E}_\epsilon \max_\theta\langle \epsilon, \theta\rangle \\
&= \sqrt{\frac{2}{\pi}}R_n(T)
\end{aligned}$$

An example set for tightness is $T = \{-1, 1\}^n$. Then, $R_n(T) = \frac{2}{n}$ and $G_n(T) = \frac{2}{n}\sqrt{\frac{2}{\pi}}$.

(b) It is not hard to show that any (gaussian) (random) vector u of can be decomposed into a sum of $\lfloor \max |u| \rfloor + 1$ vectors where the last one has entries with entries whose absolute values are at most 1, and the other ones have entries of absolute value exactly 1. In fact, let's append infinite number of zero vectors into the decomposition. Using this,

$$
\begin{aligned}
G_n(T) &= \frac{1}{n}\mathbb{E}_\omega \sup_\theta \langle \omega, \theta\rangle \\
&= \frac{1}{n}\mathbb{E}_\omega \sup_\theta \langle \sum_{j=1}^{\infty} v_j, \theta\rangle \text{ using the aforementioned decomposition} \\
&= \frac{1}{n}\mathbb{E}_\omega (\sup_\theta \sum_{j=1}^{\infty} 1_{\max|\omega|\geq j}\langle v_j, \theta\rangle + \sup_\theta\langle r, \theta\rangle) \text{ where } r \text{ is the aforementioned remainder vector} \\
&\leq \frac{1}{n}\mathbb{E}_\omega (\sum_{j=1}^{\infty} \sup_\theta 1_{\max|\omega|\geq j}\langle v_j, \theta\rangle + \sup_\theta\langle r, \theta\rangle) \\
&= \frac{1}{n}\sum_{j=1}^{\infty}\mathbb{E}_\omega (\sup_\theta 1_{\max|\omega|\geq j}\langle v_j, \theta\rangle) + \frac{1}{n}\mathbb{E}_\omega(\sup_\theta\langle r, \theta\rangle) \\
&= \frac{1}{n}\sum_{j=1}^{\infty}\mathbb{E}_\omega (1_{\max|\omega|\geq j}\sup_\theta\langle v_j, \theta\rangle) + \frac{1}{n}\mathbb{E}_\omega(\sup_\theta\langle r, \theta\rangle) \\
&= \frac{1}{n}\sum_{j=1}^{\infty}\mathbb{E}_\omega (1_{\max|\omega|\geq j})\mathbb{E}_\omega(\sup_\theta\langle v_j, \theta\rangle) + \frac{1}{n}\mathbb{E}_\omega(\sup_\theta\langle r, \theta\rangle) \text{ using distributional symmetry of } v_j \\
&= \frac{1}{n}\sum_{j=1}^{\infty}P(\max|\omega|\geq j)\mathbb{E}_\omega(\sup_\theta\langle v_j, \theta\rangle) + \frac{1}{n}\mathbb{E}_\omega(\sup_\theta\langle r, \theta\rangle) \\
&= \frac{1}{n}\sum_{j=1}^{\infty}P(\max|\omega|\geq j)\mathbb{E}_\epsilon(\sup_\theta\langle \epsilon, \theta\rangle) + \frac{1}{n}\mathbb{E}_\omega(\sup_\theta\langle r, \theta\rangle) \text{ using distributional symmetry of } v_j \\
&= \sum_{j=1}^{\infty}P(\max|\omega|\geq j)R_n(T) + \frac{1}{n}\mathbb{E}_\omega(\sup_\theta\langle r, \theta\rangle) \\
&= R_n(T)\sum_{j=1}^{\infty}P(\max|\omega|\geq j) + \frac{1}{n}\mathbb{E}_\omega(\sup_\theta\langle r, \theta\rangle) \\
&= R_n(T)\mathbb{E}_\omega(\max|\omega|) + \frac{1}{n}\mathbb{E}_\omega(\sup_\theta\langle r, \theta\rangle) \\
&= 2R_n(T)\sqrt{\log n} + \frac{1}{n}\mathbb{E}_\omega(\sup_\theta\langle r, \theta\rangle) \text{ using Homework 1/3b)} \\
&= 2R_n(T)\sqrt{\log n} + \frac{1}{n}\mathbb{E}_r(\sup_\theta\langle r, \theta\rangle) \\
&\leq 2R_n(T)\sqrt{\log n} + R_n(T) \\
&\quad \text{by further decomposing } r \text{ into infinitely many vectors containing } \pm \text{ negative powers of 2} \\
&\quad \text{and bounding similarly as the other terms (details omitted)} \\
&= O(R_n(T)\sqrt{\log n})
\end{aligned}
$$

An example set for tightness is the l1 ball. Then, $R_n(B_1^n) = \mathbb{E}[||\epsilon||_\infty] = 1$ using Holder's inequality, and $G_n(B_1^n) = \mathbb{E}[||\omega||_\infty] = O(\log n)$ using the proof presented in http://www.gautamkamath.com/writings/gaussian_max.pdf.

3. Note that "A simple geometric argument..." claimed in the question does not work per se for $\gamma \leq 0$, so in what follows I assume $\gamma > 0$, i.e. a perfect split always exists.

   (a)

$$
\begin{aligned}
P(y\langle\hat{\theta}, x\rangle < 0) &= P(y\langle[r, \gamma\tilde{\theta}], [yr, \tilde{x}]\rangle < 0) \\
&= P(y^2 r^2 + y\gamma\langle\tilde{\theta}, \tilde{x}\rangle < 0) \\
&= P(r^2 + y\gamma\langle\tilde{\theta}, \tilde{x}\rangle < 0) \\
&= \frac{1}{2}P(r^2 + \gamma\langle\tilde{\theta}, \tilde{x}\rangle < 0) + \frac{1}{2}P(r^2 - \gamma\langle\tilde{\theta}, \tilde{x}\rangle < 0) \text{ by conditioning on } y \\
&= \frac{1}{2}(P(\langle\tilde{\theta}, \tilde{x}\rangle < -\frac{r^2}{\gamma}) + P(\langle\tilde{\theta}, \tilde{x}\rangle > \frac{r^2}{\gamma})) \\
&= \frac{1}{2}P(|\langle\tilde{\theta}, \tilde{x}\rangle| > \frac{r^2}{\gamma}) \\
&= \frac{1}{2}P_{Z\sim N(0,1)}(|Z| > \frac{r^2}{\gamma}) \\
&\text{because } \tilde{x} \sim N(0, I),\ ||\tilde{\theta}||_2 = 1, \text{ so } \langle\tilde{\theta}, \tilde{x}\rangle \text{ is normal with } E[\langle\tilde{\theta}, \tilde{x}\rangle] = 0 \text{ and } V[\langle\tilde{\theta}, \tilde{x}\rangle] = ||\tilde{\theta}||_2 = 1 \\
&= P(Z > \frac{r^2}{\gamma}) \\
&= P(Z < -\frac{r^2}{\gamma}) \\
&= \text{CDF}_{N(0,1)}(-\frac{r^2}{\gamma})
\end{aligned}
$$

   (b)

$$
\begin{aligned}
\gamma &= \max_{\theta} \min_{x,y} y\frac{\langle\theta, x_{2:d}\rangle}{||\theta||_2} \\
&\leq \max_{\theta} \min_{x} |\frac{\langle\theta, x_{2:d}\rangle}{||\theta||_2}| \\
&\leq \max_{\theta} \frac{||\frac{\tilde{X}\theta}{||\theta||_2}||_2}{\sqrt{n}} \text{ by a simple inequality between minimum and l2 norm} \\
&= \frac{1}{\sqrt{n}} \max_{\theta} \frac{||\tilde{X}\theta||_2}{||\theta||_2} \\
&= \frac{1}{\sqrt{n}} \max_{\theta} \frac{\sqrt{\theta^T\tilde{X}^T\tilde{X}\theta}}{||\theta||_2} \\
&\leq \frac{1}{\sqrt{n}} \max_{\theta} \frac{s_{max}(\tilde{X})||\theta||_2}{||\theta||_2} \text{ using exactly the same algebra as in 1/b) bounding the expectation} \\
&= \frac{1}{\sqrt{n}} \max_{\theta} s_{max}(\tilde{X}) \\
&= \frac{s_{max}(\tilde{X})}{\sqrt{n}}
\end{aligned}
$$

(c) i.

$$
\begin{aligned}
\mathbb{E}[(X_{u,v} - X_{u',v'})^2] &= \mathbb{E}[(\sum_{ij} X_{ij}(u_j v_i - u'_j v'_i))^2] \\
&= \mathbb{E}[\sum_{ijkl} X_{ij}X_{kl}(u_j v_i - u'_j v'_i)(u_k v_l - u'_k v'_l)] \\
&= \sum_{ijkl} \mathbb{E}[X_{ij}X_{kl}](u_j v_i - u'_j v'_i)(u_k v_l - u'_k v'_l) \\
&= \sum_{ij} \mathbb{E}[X_{ij}^2](u_j v_i - u'_j v'_i)^2 \\
&\text{due to the independence among the entries of } X \text{ and } \mathbb{E}[X_{ij}] = 0 \\
&= \sum_{ij}(u_j v_i - u'_j v'_i)^2 \text{ as } X_{ij}^2 \text{ are chi-squared} \\
&= \sum_i {u_i}^2 \sum_i {v_i}^2 + \sum_i {u'_i}^2 \sum_i {v'_i}^2 - 2\sum_i u_j u'_j \sum_i v_i v'_i \\
&= ||u||_2^2 ||v||_2^2 + ||u'||_2^2 ||v'||_2^2 - 2\langle u, u'\rangle\langle v, v'\rangle \\
&= 2 - 2\langle u, u'\rangle\langle v, v'\rangle \text{ due to the norm assumption on the vectors}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[(Y_{u,v} - Y_{u',v'})^2] &= \mathbb{E}[(\sum_i g_i(u_i - u'_i) + h_i(v_i - v'_i))^2] \\
&= \sum_i \mathbb{E}[g_i^2](u_i - u'_i)^2 + \mathbb{E}[h_i^2](v_i - v'_i)^2 \\
&\text{due to independence between } g \text{ and } h \text{ and among different entries of } g,\, h \text{ and } \mathbb{E}[g_i] = \mathbb{E}[h_i] = 0 \\
&= \sum_i (u_i - u'_i)^2 + (v_i - v'_i)^2 \\
&= ||u||_2^2 + ||u'||_2^2 + ||v||_2^2 + ||v'||_2^2 - 2\langle u, u'\rangle - 2\langle v, v'\rangle \\
&= 4 - 2\langle u, u'\rangle - 2\langle v, v'\rangle
\end{aligned}
$$

Combining the two calculations and rearranging, it only remains to show that

$$\langle u, u'\rangle + \langle v, v'\rangle < 1 + \langle u, u'\rangle\langle v, v'\rangle$$

this is equivalent to

$$\langle u, u'\rangle(1 - \langle v, v'\rangle) < 1 - \langle v, v'\rangle$$

since $\langle v, v'\rangle \leq ||v||_2||v'||_2 \leq 1$, we can divide by $1 - \langle v, v'\rangle$ getting

$$\langle u, u'\rangle < 1$$

which holds by Cauchy-Schwartz and norm bound.

ii.

$$
\begin{aligned}
\mathbb{E}[s_{max}(X)] &= \mathbb{E}[\max_u \max_v \langle Xu, v\rangle] \\
&= \mathbb{E}[\sup_{(u,v)\in S^{d-1}\times S^{n-1}} \langle Xu, v\rangle] \\
&\leq \mathbb{E}[\sup_{(u,v)\in S^{d-1}\times S^{n-1}} \langle g, u\rangle + \langle h, v\rangle] \text{ using Slepian's inequality} \\
&\leq \mathbb{E}\sup_u \langle g, u\rangle + \mathbb{E}\sup_v \langle h, v\rangle \\
&\leq \mathbb{E}\sup_u ||g||_2||u||_2 + \mathbb{E}\sup_v ||h||_2||v||_2 \\
&= \mathbb{E}||g||_2 + \mathbb{E}||h||_2 \\
&\leq \sqrt{n} + \sqrt{d} \text{ using Jensen as in question 1/b}
\end{aligned}
$$

(d) First let's show that s_{max} is 1-Lipschitz.

$$\begin{aligned}
|s_{max}(X_1) - s_{max}(X_2)| &= |\sup_{uv}\langle X_1 u, v\rangle - \sup uv\langle X_2 u, v\rangle| \\
&\leq |\sup_{uv}\langle X_1 u, v\rangle - \langle X_2 u, v\rangle| \\
&= |\sup_{uv}\langle (X_1 - X_2)u, v\rangle| \\
&\leq |\sup_{uv}||(X_1 - X_2)u||_2||v||_2| \\
&= \sup_{u}||(X_1 - X_2)u||_2 \\
&\leq \sup_{u}||(X_1 - X_2)||_2||u||_2 \\
&= ||(X_1 - X_2)||_2
\end{aligned}$$

Then,

$$\begin{aligned}
P(s_{max}(\tilde{X}) \leq \sqrt{d} + \sqrt{n} + t) &= 1 - P(s_{max}(\tilde{X}) \geq \sqrt{d} + \sqrt{n} + t) \\
&\geq 1 - P(s_{max}(\tilde{X}) \geq \sqrt{d-1} + \sqrt{n} + t) \text{ note that} \\
&\geq 1 - P(s_{max}(\tilde{X}) \geq \mathbb{E}[s_{max}(\tilde{X})] + t) \text{ using part c)} \\
&\geq 1 - 2e^{-\frac{t^2}{2}} \text{ using Theorem 2.26 from MW and the Lipschitzness of } s_{max}
\end{aligned}$$

4. Choice 1

(a) Intuitively, there is no reason to assume that the generalisation error should depend on D. After all, if all the xs are scaled by some factor c, $\langle w, x\rangle$ is also scaled by c but the sign is unchanged. Hence, the predictions should not be affected. In my understanding, the dependence comes in via the notion of Rademacher complexity used in our bound, which is unlike the actual predictions, not invariant to scaling.

(b) One easy way to get around this, is to change the bound on $||w||_2$ to $\frac{B}{D}$ instead of B. This also does not affect the predictions as it is a scaling factor, but it would make the D cancel out from the formula. Note however, that while it looks great, it changes the scale of γ in the sense that for the same γ (e.g. in the Rademacher term), $R_n^\gamma(f)$ will be larger, as the margins are smaller due to the smaller ws. So in some sense, we achieved something in the formula, but we did not get a smaller bound. Another alternative, which is probably more reasonable is to restrict not the l2-norm of w but rather the output of the function directly. There are 2 merits in doing so – it can provide a way to bound $\langle w, x\rangle$ without regards to $||x||_2$, and also note that original loss function never, and the ramp loss also cares about the exact magnitude of the output only in the range $0 - \gamma$.