A Details of Uniform Sampling Lower Bound

In this section we repeat the statement of Theorem 5 and give a full proof.

Theorem 5. *Assume that the loss function is either logistic regression or SVM, and the regularizer is the 2-norm squared. Let $\epsilon, \gamma \in (0,1)$ be arbitrary. For all sufficiently large n, there exist an instance I_n of n points such that with probability at least $1 - 1/n^{\gamma/2}$ it will be the case that for a uniform sample C of $c = n^{1-\gamma}/\lambda$ points, there is no weighting U that will result in an ϵ-coreset.*

Proof. The instance I_n consists of points located on the real line, so the dimension $d = 1$. A collection A of $n - (\lambda n^{\gamma/2})$ points is located at $+1$, and the remaining $\lambda n^{\gamma/2}$ points are located at -1, call this collection of points B. All points are labeled $+1$. Note $R = 1$.

Let C be the random sample of c points, and U an arbitrary weighting of the points in C. Note that U may depend on the instantiation of C. Our goal is to show that (C, U) is not an ϵ-coreset. Our proof strategy is to first show that its likely that C contains only points from A. Then we want to show that, conditioned on $C \subseteq A$, that C can not be a core set for any possible weighting. We accomplish this by showing that the limit as n approaches infinity of the lefthand size of the definition of coreset in equation (2) must be 1.

We now show that one can use a standard union bound to establish that it is likely that $C \subseteq A$. To accomplish this let E_i be the probability that the the i^{th} point selected to be in C is not in A.

$$\Pr[C \subseteq A] = 1 - \Pr\left[\vee_{i\in C} E_i\right] \geq 1 - |C|\frac{|B|}{n}$$
$$= 1 - \frac{n^{1-\gamma}}{\lambda}\frac{\lambda n^{\gamma/2}}{n} = 1 - \frac{1}{n^{\gamma/2}}$$

Now we show if $C \subseteq A$ and n is large enough, then (C, U) cannot be an ϵ-coreset for any collection U of weights. To accomplish this consider the the hypothesis $\beta_0 = n^{\gamma/4}$. From the definition of coreset, it is sufficient to show that $H(\beta_0)$, defined as,

$$H(\beta_0) = \frac{|\sum_{i\in P} f_i(\beta_0) - \sum_{i\in C} u_i f_i(\beta_0)|}{\sum_{i\in P} f_i(\beta_0)} \quad (4)$$

is greater than ϵ. We accomplish this by showing that the limit as n goes to infinity of $H(\beta_0)$ is 1. Applying Condition 1 we can conclude that

$$H(\beta_0) \geq \frac{|\sum_{i\in P} \ell_i(\beta_0) - \sum_{i\in C} u_i \ell_i(\beta_0)| - \epsilon\lambda \|\beta_0\|_2^2}{\sum_{i\in P} \ell_i(\beta_0) + \lambda \|\beta_0\|_2^2} \quad (5)$$

Then, using the fact that A and B is a partition of the points and $C \subseteq A$ we can conclude that

$$H(\beta_0) \geq \frac{|\sum_{i\in A} \ell_i(\beta_0) + \sum_{i\in B} \ell_i(\beta_0) - \sum_{i\in C} u_i \ell_i(\beta_0)| - \epsilon\lambda \|\beta_0\|_2^2}{\sum_{i\in A} \ell_i(\beta_0) + \sum_{i\in B} \ell_i(\beta_0) + \lambda \|\beta_0\|_2^2}$$
$$= \frac{\left|\frac{\sum_{i\in A} \ell_i(\beta_0)}{\sum_{i\in B} \ell_i(\beta_0)} + 1 - \frac{\sum_{i\in C} u_i \ell_i(\beta_0)}{\sum_{i\in B} \ell_i(\beta_0)}\right| - \frac{\epsilon\lambda\|\beta_0\|_2^2}{\sum_{i\in B} \ell_i(\beta_0)}}{\frac{\sum_{i\in A} \ell_i(\beta_0)}{\sum_{i\in B} \ell_i(\beta_0)} + 1 + \frac{\lambda\|\beta_0\|_2^2}{\sum_{i\in B} \ell_i(\beta_0)}} \quad (6)$$

We now need to bound various terms in equation (6). Let us first consider logistic regression. Note that

$$\sum_{i\in B} \ell_i(\beta_0) = |B|\log(1 + \exp(n^{\gamma/4})) \geq |B|n^{\gamma/4} = \lambda n^{3\gamma/4} \quad (7)$$

Therefore,

$$\lim_{n\to\infty} \frac{\lambda \|\beta_0\|_2^2}{\sum_{i\in B} \ell_i(\beta_0)} \leq \lim_{n\to\infty} \frac{\lambda n^{\gamma/2}}{\lambda n^{3\gamma/4}} = 0 \quad (8)$$

Also note that

$$\lim_{n\to\infty} \sum_{i\in A} \ell_i(\beta_0) \quad (9)$$
$$= \lim_{n\to\infty} |A|\log(1 + \exp(-n^{\gamma/4})) \quad (10)$$
$$\leq \lim_{n\to\infty} n\exp(-n^{\gamma/4}) = 0 \quad (11)$$

Finally, by Observation 1, we have,

$$\lim_{n\to\infty} \sum_{i\in C} u_i \ell_i(\beta_0) \leq \lim_{n\to\infty} (1+\epsilon) n \exp(-n^{\gamma/4}) = 0 \quad (12)$$

Combining equations (7), (8), (9), and (12), we the expression in equation (6) converges to 1 as $n \to \infty$. Thus for sufficiently large n, $H(\beta_0) > \epsilon$and thus (C, U) is not an ϵ-coreset.

We now need to bound various terms in equation (6) for SVM. First note that

$$\sum_{i\in B} \ell_i(\beta_0) = |B|(1 + n^{\gamma/4}) \geq |B|n^{\gamma/4} = \lambda n^{3\gamma/4} \quad (13)$$

Therefore,

$$\lim_{n\to\infty} \frac{\lambda \|\beta_0\|_2^2}{\sum_{i\in B} \ell_i(\beta_0)} \leq \lim_{n\to\infty} \frac{\lambda n^{\gamma/2}}{\lambda n^{3\gamma/4}} = 0 \quad (14)$$

Also note that

$$\lim_{n\to\infty} \sum_{i\in A} \ell_i(\beta_0) = \lim_{n\to\infty} |A| \max(0, 1 - n^{\gamma/4}) = 0 \quad (15)$$

Finally, by Observation 1, we have that:

$$\lim_{n\to\infty}\sum_{i\in C}u_i\ell_i(\beta_0)\le\lim_{n\to\infty}(1+\epsilon)n\max(0,1-n^{\gamma/4})=0 \tag{16}$$

Combining equations (13), (14), (15), and (16), we the expression in equation (6) converges to 1 as $n\to\infty$. Thus for sufficiently large n, $H(\beta_0)>\epsilon$and thus (C,U) is not an ϵ-coreset. □

B Details of General Lower Bound on Coreset Size

B.1 Logistic Regression

Proof of Observation 9:

Proof. It is well known that

$$d_i=\frac{|(\beta_x,\beta_y)\cdot x_i+\beta_z|}{\sqrt{\beta_x^2+\beta_y^2}}$$

Therefore,

$$|(\beta_x,\beta_y)\cdot x_i+\beta_z|=d_i\sqrt{\beta_x^2+\beta_y^2}\le\|\beta_A\|\,d_i$$

Now we need to show $\|\beta_A\|\,d_i/2\le|(\beta_x,\beta_y)\cdot x_i+\beta_z|$. Note that there are two points (points adjacent to A) $x_j=(a',b')$ for which $(\beta_x,\beta_y)\cdot x_j+\beta_z=0$. Consider one of them. We have:

$$\begin{aligned}0&=\beta_x a'+\beta_y b'+\beta_z\\&\ge\beta_z-|\beta_x a'+\beta_y b'|\\&\ge\beta_z-\sqrt{\beta_x^2+\beta_y^2}\sqrt{a'^2+b'^2}\end{aligned}$$

Since the points are over a circle of size 1 we have $\sqrt{a'^2+b'^2}=1$. Therefore,

$$\beta_x^2+\beta_y^2\ge\beta_z^2$$

So we can conclude:

$$|(\beta_x,\beta_y)\cdot x_i+\beta_z|=d_i\sqrt{\beta_x^2+\beta_y^2}\ge\frac{d_i}{\sqrt2}\|\beta_A\|\ge\frac{d_i}{2}\|\beta_A\|$$

□

Remaining Proof of Lemma 8:

Proof. Using the Taylor expansion of $\cos(x)=\sum_{i=0}^{\infty}(-1)^i\frac{x^{2i}}{(2i)!}=1-\frac{x^2}{2!}+\frac{x^4}{4!}-\ldots$, we have $\cos(\frac{\pi}{4k})-\cos(\frac{\pi}{2k})\ge\frac12\big((\frac{\pi}{2k})^2-(\frac{\pi}{4k})^2\big)-O(\frac{1}{k^4})=(\frac{3\pi^2}{32k^2})-O(\frac{1}{k^4})$. Plugging this inequality, we derive

$$\begin{aligned}&\frac{\sum_{x_i\in C}u_i\ell_i(\beta_A)}{\sum_{x_i\in P}\ell_i(\beta_A)}\\&\le\frac{\sum_{x_i\in C}u_i\exp(-\frac{\|\beta_A\|}{2}((\frac{3\pi^2}{32k^2})-O(\frac{1}{k^4})))}{\sum_{x_i\in A}\ell_i(\beta_A)}\\&=\frac{\sum_{x_i\in C}u_i\exp(-\frac{n^{2/5}}{2\sqrt{c}\lambda^{2/5}}(\frac{3\pi^2\lambda^{2/5}}{32c^2n^{2/5-2\gamma}}-O(\frac{\lambda^{4/5}}{n^{4/5-4\gamma}})))}{\sum_{x_i\in A}\ell_i(\beta_A)}\\&=\frac{\sum_{x_i\in C}u_i\exp(-\alpha n^{2\gamma}+O(\frac{\lambda^{2/5}}{n^{2/5-4\gamma}}))}{\sum_{x_i\in A}\ell_i(\beta_A)},\end{aligned}$$

where $\alpha>0$ is a constant. Since all the points in A are miss-classified, we have $\ell_i(\beta_A)\ge\log2$ for all of them. Using this fact and Observation 1, we have:

$$\frac{\sum_{x_i\in C}u_i\ell_i(\beta_A)}{\sum_{x_i\in P}\ell_i(\beta_A)}\le\frac{(1+\epsilon)n\exp(-\alpha n^{2\gamma}+O(\frac{\lambda^{2/5}}{n^{2/5-4\gamma}}))}{\frac{n}{4k}\log2}$$

Finally, using the fact that $k=c\frac{n^{1/5-\gamma}}{\lambda^{1/5}}$ and taking the limit, we conclude:

$$\begin{aligned}&\lim_{n\to\infty}\frac{\sum_{x_i\in C}u_i\ell_i(\beta_A)}{\sum_{x_i\in P}\ell_i(\beta_A)}\\&\le\lim_{n\to\infty}\frac{4k(1+\epsilon)\exp(-\alpha n^{2\gamma}+O(\frac{\lambda^{2/5}}{n^{2/5-4\gamma}}))}{\log2}=0\end{aligned}$$

□

B.2 SVM

For the sake of contradiction, suppose an ϵ-coreset (C,u) of size k, as stated in Theorem 6, exists for the circle instance. We fix A to be a chunk. Similar to logistic regression, we set β_A as the parameters of the linear SVM that separates A from P/A such that the model predicts A incorrectly and predicts the points P/A as positive correctly and $\|\beta_A\|_2=\sqrt{\frac{n^{1-\gamma}}{k\lambda}}=\frac{n^{2/5}}{\sqrt{c}\lambda^{2/5}}$.

Our goal is to show Eqn. (2) tends to 1 as n grows to infinity. We can break the cost function of Linear SVM into two parts:

$$F_{P,1}(\beta_A):=\sum_{x_i\in P}\ell_i(\beta_A)+2\lambda\|\beta_A\|_2^2$$

where $\ell_i(\beta_A)=\max(1-\beta_A x_i y_i,0)=\max(1-((\beta_x,\beta_y)\cdot x_i+\beta_z)y_i,0)$. Then, we determine the limit of the following quantities as n grows to infinity.

Lemma 11. *For the circle instance P, if (C,u) is an ϵ-coreset of P with size $k=c\frac{n^{1/5-\gamma}}{\lambda^{1/5}}$ for linear SVM, and A is a chunk, then we have,*

1. $\lim_{n\to\infty} \frac{\lambda\|\beta_A\|_2^2}{\sum_{x_i\in P} \ell_i(\beta_A)} = 0$;

2. $\lim_{n\to\infty} \frac{\sum_{x_i\in C} u_i\ell_i(\beta_A)}{\sum_{x_i\in P} \ell_i(\beta_A)} = 0$.

Using this lemma, which we will prove soon, we can prove Theorem 6 for the linear SVM: The definition of coreset allows us to choose any β, so we can set $\beta = \beta_A$ for a chunk A. Then, by Observation 1, Eqn. (2) simplifies to:

$$\begin{aligned}
&\frac{|\sum_{x_i\in X} f_i(\beta_A) - \sum_{x_i\in C} u_i f_i(\beta_A)|}{\sum_{x_i\in X} f_i(\beta_A)} \\
&\geq \frac{|\sum_{x_i\in X} \ell_i(\beta_A) - \sum_{x_i\in C} u_i\ell_i(\beta_A)| - 2\epsilon\lambda\|\beta_A\|_2^2}{\sum_{x_i\in X} \ell_i(\beta_A) + \lambda\|\beta_A\|_2^2} \\
&= \frac{|1 - \frac{\sum_{x_i\in C} u_i\ell_i(\beta_A)}{\sum_{x_i\in X} \ell_i(\beta_A)}| - \frac{2\epsilon\lambda\|\beta_A\|_2^2}{\sum_{x_i\in X} \ell_i(\beta_A)}}{1 + \frac{\lambda\|\beta_A\|_2^2}{\sum_{x_i\in X} \ell_i(\beta_A)}},
\end{aligned}$$

which tends to 1 as $n \to \infty$ by Lemma 11. This implies that (C, u) is not an ϵ-coreset for the circle instance, which is a contradiction. This completes the proof of Theorem 6 for SVM.

Proof of Lemma 11

The remainder of this section is devoted to proving Lemma 11. The proof is very similar to the proof of Lemma 7 and 8.

Proof. **of Claim 1 in Lemma 11** We know for all points in A, $\ell_i(\beta_A) \geq 1$ this is because all of them have been incorrectly classified. We also know that since A is a chunk, $|A| = \frac{n}{4k}$.

Therefore

$$\sum_{x_i\in A} \ell_i(\beta_A) \geq \frac{n}{4k}$$

We also know $\|\beta_A\|_2 = \sqrt{\frac{n^{1-\gamma}}{k\lambda}}$, so we can conclude

$$\frac{\lambda\|\beta_A\|_2^2}{\sum_{x_i\in X} \ell_i(\beta_A)} \leq \frac{\lambda\|\beta_A\|_2^2}{\sum_{x_i\in A} \ell_i(\beta_A)} \leq \frac{\lambda\|\beta_A\|_2^2}{\frac{n}{4k}} = \frac{\lambda\frac{n^{1-\gamma}}{k\lambda}}{\frac{n}{4k}} = \frac{4}{n^\gamma}$$

The lemma follows by taking the limit of the above inequality. ☐

Proof. **of Claim 2 in Lemma 11**. Using Observation 9 and the fact that all the points in the coreset are predicted correctly by β_A we have:

$$\frac{\sum_{x_i\in C} u_i\ell_i(\beta_A)}{\sum_{x_i\in X} \ell_i(\beta_A)} \leq \frac{\sum_{x_i\in C} u_i \max\left(0, 1 - \frac{\|\beta_A\| d_i}{2}\right)}{\sum_{x_i\in A} \ell_i(\beta_A)}$$

Then, by Observation 10 we have

$$\begin{aligned}
&\frac{\sum_{x_i\in C} u_i\ell_i(\beta_A)}{\sum_{x_i\in X} \ell_i(\beta_A)} \\
&\leq \frac{\sum_{x_i\in C} u_i \max\left(0, 1 - \frac{\|\beta_A\|}{2}(\cos\theta - \cos\theta_i)\right)}{\sum_{x_i\in A} \ell_i(\beta_A)}
\end{aligned}$$

By definition of chunk, we know all the points in C are at least $\frac{n}{4k}$ away from the center of A, which means the closest point in C to chunk A is at least $\frac{n}{8k}$ points away, we have $\theta_i \geq \theta + \frac{2\pi}{n}\frac{n}{8k} = \frac{\pi}{2k}$. Therefore,

$$\begin{aligned}
&\frac{\sum_{x_i\in C} u_i\ell_i(\beta_A)}{\sum_{x_i\in X} \ell_i(\beta_A)} \\
&\leq \frac{\sum_{x_i\in C} u_i \max\left(0, 1 - \frac{\|\beta_A\|}{2}(\cos\frac{\pi}{4k} - \cos\frac{\pi}{2k})\right)}{\sum_{x_i\in A} \ell_i(\beta_A)}
\end{aligned}$$

Using the Taylor expansion of $\cos(x) = \sum_{i=0}^{\infty}(-1)^i \frac{x^{2i}}{(2i)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \ldots$, we have,

$$\begin{aligned}
&\cos(\frac{\pi}{4k}) - \cos(\frac{\pi}{2k}) \\
&\geq \frac{1}{2}\left((\frac{\pi}{2k})^2 - (\frac{\pi}{4k})^2\right) - O(\frac{1}{k^4}) = (\frac{3\pi^2}{32k^2}) - O(\frac{1}{k^4})
\end{aligned}$$

Therefore, we derive,

$$\begin{aligned}
&\frac{\sum_{x_i\in C} u_i\ell_i(\beta_A)}{\sum_{x_i\in X} \ell_i(\beta_A)} \\
&\leq \frac{\sum_{x_i\in C} u_i \max\left(0, 1 - \frac{\|\beta_A\|}{2}((\frac{3\pi^2}{32k^2}) - O(\frac{1}{k^4}))\right)}{\sum_{x_i\in A} \ell_i(\beta_A)} \\
&= \frac{\sum_{x_i\in C} u_i \max\left(0, 1 - \frac{n^{2/5}}{2c\lambda^{2/5}}(\frac{3\pi^2\lambda^{2/5}}{32cn^{2/5-2\gamma}} - O(\frac{\lambda^{4/5}}{n^{4/5-4\gamma}}))\right)}{\sum_{x_i\in A} \ell_i(\beta_A)} \\
&= \frac{\sum_{x_i\in C} u_i \max\left(0, 1 - \alpha n^{2\gamma} + O(\frac{\lambda^{2/5}}{n^{2/5-4\gamma}})\right)}{\sum_{x_i\in A} \ell_i(\beta_A)}
\end{aligned}$$

For large enough n, we have $\max\left(0, 1 - \alpha n^{2\gamma} + O(\frac{\lambda^{2/5}}{n^{2/5-4\gamma}})\right) = 0$. Therefore, by taking the limit we have:

$$\lim_{n\to\infty} \frac{\sum_{x_i\in C} u_i\ell_i(\beta_A)}{\sum_{x_i\in X} \ell_i(\beta_A)}$$

$$\leq \lim_{n\to\infty} \frac{\sum\limits_{x_i\in C} u_i \max\left(0, 1-\alpha n^{2\gamma} + O(\frac{\lambda^{2/5}}{n^{2/5-4\gamma}})\right)}{\sum\limits_{x_i\in A} \ell_i(\beta_A)} = 0$$

□