1) Fix $\varepsilon > 0$ | Math 597 A1 J. Pearce 260672004

Let R_s be the smallest hyper-rectange in $\mathbb{R}^n$ consistent with data

$\mathbb{P}(R) < \varepsilon \Rightarrow \mathbb{P}(R_s) < \varepsilon \Rightarrow R(R_s) < \varepsilon$ Done

$\Rightarrow$ Assume $\mathbb{P}(R) > \varepsilon$

Define $2n$ hyper rectangles $r_1, r_2, \ldots r_{2n}$. To create r_i, start with R and then decrease the size by moving a hyperplane of R (i.e. one of $\{a_1, b_1, a_2, b_2, \ldots, a_n, b_n\}$) as much as possible while maintaing $\mathbb{P}(r_i) \geq \frac{\varepsilon}{2n}$.

$\Rightarrow \mathbb{P}(r_i) \geq \frac{\varepsilon}{2n} \quad \forall i$

$\Rightarrow \mathbb{P}(R) - \mathbb{P}(\bigcup_{i=1}^{n} r_i) < \varepsilon$

$\Rightarrow$ If $R(R_s') > \varepsilon$, then R_s must miss at least one r_i

$$\Rightarrow \mathbb{P}_{S\sim D^m}[R(R_s') > \varepsilon] \leq \mathbb{P}_{S\sim D^m}\left[\bigcup_{i=1}^{2n} \{R_s' \cap r_i = \emptyset\}\right]$$

$$\leq \sum_{i=1}^{2n} \mathbb{P}_{S\sim D^m}[R_s \cap r_i = \emptyset] \quad \text{(by union bound)}$$

$$\leq 2n\left(1 - \frac{\varepsilon}{2n}\right)^m \quad \left(\mathbb{P}(r_i) \geq \frac{\varepsilon}{2n}\right)$$

$$\leq 2n \exp\left(-\frac{m\varepsilon}{2n}\right) \quad (1 - x \leq e^{-x})$$

$$\boxed{\Rightarrow \mathbb{P}_{S\sim D^m}[R(R_s) > \varepsilon] \leq 2n \exp\left(\frac{-m\varepsilon}{2n}\right)}$$

Now for any $\delta > 0$, we want RHS $\leq 1-\delta$

Solve for m:

$$2n \exp\left(\frac{-m\varepsilon}{2n}\right) = \delta$$

$$\frac{m\varepsilon}{2n} = -\log\left(\frac{\delta}{2n}\right)$$

$$m = \frac{2n}{\varepsilon}\log\left(\frac{2n}{\delta}\right)$$

$$\boxed{\Rightarrow \text{ when } m \geq \frac{2n}{\varepsilon}\log\left(\frac{2n}{\delta}\right) \text{ we get } \mathbb{P}[R(R_s) < \varepsilon] \geq 1-\delta}$$

2) Algorithm : for training sample S, return hypothesis I_S : let $[a', b'] \subseteq [a,b]$, $[c', d'] \subseteq [c,d]$

If there are 2 seperate sequences of positive labels return ~~the~~ $[a', b'] \cup [c', d']$ where $[a', b']$ is the smallest interval containing the first sequence of ~~the~~ positive points, and $[c', d']$ the smallest interval for the second sequence

Else return $[a', d']$ the smallest interval containing all the positive points where $[a', d'] = [a', b'] \cup [c', d']$ with $c' = b'$

Let $\varepsilon > 0$, $I \in C_2$

$\mathbb{P}(I) < \varepsilon \implies \mathbb{P}(I_S) < \varepsilon \implies R(I_S) < \varepsilon$ done

$\implies \mathbb{P}(I) > \varepsilon \longrightarrow$ assume $\mathbb{P}([a,b]) > \varepsilon/3$, $\mathbb{P}([c,d]) > \varepsilon/3$

Define 4 intervals r_1, r_2, r_3, r_4. r_1 is an interval of the form $[a, b']$ where $b' \leq b$ and b' is as small as possible while maintaining $\mathbb{P}([a,b']) \geq \frac{\varepsilon}{6}$. r_2, r_3, r_4 are constructed in the same manner.

$\implies \mathbb{P}(r_i) \geq \frac{\varepsilon}{6} \quad \forall i \in \{1,2,3,4\}$

If $R(I_S) > \varepsilon$, I_S misses at least one r_i or $\mathbb{P}[(b,c)] > \frac{\varepsilon}{3}$ and no sample point is in (b,c)

$$\Longrightarrow \underset{S \sim D^m}{\mathbb{P}}\left[R(I_S) > \varepsilon\right] \leq \underset{S \sim D^m}{\mathbb{P}}\left[\bigcup_{i=1}^{4} I_S \cap r_i = \emptyset\right] + \left(1 - \frac{\varepsilon}{3}\right)^m$$

$$\leq \sum_{i=1}^{4} \underset{S \sim D^m}{\mathbb{P}}\left[I_S \cap r_i = \emptyset\right] + \exp\left(\frac{-m\varepsilon}{3}\right)$$ by union bound and $1-x \leq e$

$$\leq 4\left(1-\frac{\varepsilon}{6}\right)^m + \exp\left(\frac{-m\varepsilon}{3}\right) \qquad \mathbb{P}(r_i) \geq \frac{\varepsilon}{6}$$

$$\leq 4\exp\left(\frac{-m\varepsilon}{6}\right) + \exp\left(\frac{-m\varepsilon}{3}\right) \qquad 1-x < e^{-x}$$

$$\leq 5\exp\left(\frac{-m\varepsilon}{6}\right)$$

$$\boxed{\Rightarrow \mathbb{P}_{S\sim D^m}\left[R(I_S) > \varepsilon\right] \leq 5\exp\left(\frac{-m\varepsilon}{6}\right)}$$

Solve for m $\quad 5\exp\left(\frac{-m\varepsilon}{6}\right) = \delta$

$$\frac{m\varepsilon}{6} = -\log\left(\frac{\delta}{5}\right)$$

$$m = \frac{6}{\varepsilon}\log\left(\frac{5}{\delta}\right)$$

$$\boxed{\Rightarrow \text{when } m \geq \frac{6}{\varepsilon}\log\left(\frac{5}{\delta}\right) \text{ we get } \mathbb{P}_{S\sim D^m}\left[R(I_S) < \varepsilon\right] \geq 1-\delta}$$

case C_p

Algorithm : Suppose k seperate sequences of positive labels. return $\bigcup_{i=1}^{p} [a_i', b_i']$ where for $i=1,\ldots,k-1$ $[a_i', b_i']$ is the smallest interval containing the i^{th} sequence, and where $\bigcup_{i=k}^{p} [a_i', b_i'] = [a_k', b_p']$ is the smallest interval containing the k^{th} sequence.

Let $\varepsilon > 0$

$\mathbb{P}(I) < \varepsilon \implies \mathbb{P}(I_s) < \varepsilon \implies R(I_s) < \varepsilon$. <u>done</u>

$\implies \mathbb{P}(I) > \varepsilon$

Assume $\mathbb{P}([a_i, b_i]) > \frac{\varepsilon}{2p-1}$

Define $2p$ intervals, $r_1, r_2, \ldots r_{2p}$. These intervals are constructed in the same way as the C_2 case above, but with $\mathbb{P}([a_i, b_i]) \geq \frac{\varepsilon}{4p-2}$.

$\implies \mathbb{P}(r_i) \geq \frac{\varepsilon}{4p-2} \quad \forall i \in [2p]$

If $R(I_s) > \varepsilon$, I_s misses at least one r_i <u>or</u> $\mathbb{P}[(b_i, a_{i+1})] > \frac{\varepsilon}{2p-1}$ and no sample point is in (b_i, a_{i+1}) for some $i \in \{1, \ldots p-1\}$

$$\implies \mathbb{P}_{S\sim D^m}[R(I_s) > \varepsilon] \leq \mathbb{P}_{S\sim D^m}\left[\bigcup_{i=1}^{2p} I_s \cap r_i = \emptyset\right] + (p-1)\left(1-\frac{\varepsilon}{2p-1}\right)^m$$

$$\leq \sum_{i=1}^{2p} \mathbb{P}_{S\sim D^m}(I_s \cap r_i = \emptyset) + (p-1)\exp\left(\frac{-m\varepsilon}{2p-1}\right) \quad \left(\begin{array}{l}\text{union bou}\\ 1-x < e^{-x}\end{array}\right.$$

$$\leq 2p\left(1-\frac{\varepsilon}{4p-2}\right)^m + (p-1)\exp\left(\frac{-m\varepsilon}{2p-1}\right) \quad \left(\mathbb{P}(r_i) \geq \frac{\varepsilon}{4p-}\right.$$

$$\leq 2p\exp\left(\frac{-m\varepsilon}{4p-2}\right) + (p-1)\exp\left(\frac{-m\varepsilon}{2p-1}\right) \quad \left(1-x < e^{-}\right.$$

$$\leq 3p-1 \exp\left(\frac{-m\varepsilon}{4p-2}\right)$$

$$\boxed{\implies \mathbb{P}_{S\sim D^m}[R(I_s) > \varepsilon] \leq (3p-1)\exp\left(\frac{-m\varepsilon}{4p-2}\right)}$$

Solve for m $\quad (3p-1)\exp\left(\frac{-m\varepsilon}{4p-2}\right) = \delta$

$$\frac{m\varepsilon}{4p-2} = -\log\left(\frac{\delta}{3p-1}\right)$$

$$m = \frac{4p-2}{\varepsilon}\log\left(\frac{3p-1}{\delta}\right)$$

$\Rightarrow$ when $m \geq \frac{4p-2}{\varepsilon}\log\left(\frac{3p-1}{\delta}\right)$ we get $\underset{S\sim D^m}{\mathbb{P}}\left[R(I_S) < \varepsilon\right] \geq 1-\delta$

our algorithm's time complexity is $O(m)$

3) we can start by using the basic lemma from class since all the conditions are still met.

$$\mathbb{E}[q(x_{n+1}) \mid x_n] \le (1-\mu h_n)q(x_n) - h_n(f(x_n) - f^*) + \tfrac{1}{2}h_n^2(g_n^2 + \sigma^2)$$

$$= (1-\mu h_n)q(x_n) - h_n(f(x_n) - f^*) + \tfrac{1}{2}h_n^2 g_n^2 + \tfrac{1}{2}h_n^2\sigma^2$$

$\Rightarrow$ we must show $-h_n(f(x_n) - f^*) + \frac{1}{2}h_n^2 g_n^2 \le 0$ In order to get the desired inequality

$$\tfrac{1}{2}h_n^2 g_n^2 - h_n(f(x_n) - f^*) \le \tfrac{1}{2L}h_n g_n^2 - h_n(f(x_n) - f^*) \quad \text{by } h_n \le \tfrac{1}{L}$$

$$\le h_n(f(x_n) - f^*) - h_n(f(x_n) - f^*) \quad \text{by } (4)$$

$$= 0$$

note: where (4) is equation (4) in the SGD notes on Arxiv that is a consequence of f being L-smooth.

$$\Rightarrow \mathbb{E}[q(x_{n+1}) \mid x_n] \le (1-\mu h_n)q(x_n) + \tfrac{1}{2}h_n^2\sigma^2$$

4) a) $f(x) = \frac{1}{m} \sum_{i=1}^{m} \frac{(x - x_i)^2}{2}$

$$\nabla f(x) = \frac{1}{m} \sum_{i=1}^{m} (x - x_i)$$

$$= \frac{1}{m} \left[x - x_1 + x - x_2 + \dots + x - x_m \right]$$

$$= \frac{1}{m} \left[mx - \sum x_i \right]$$

$$= x - \frac{1}{m} \sum_{i=1}^{m} x_i$$

L-Smooth

$$\| \nabla f(x) - \nabla f(y) \| = \| x - \frac{1}{m} \sum_{1}^{m} x_i - \left(y - \frac{1}{m} \sum_{i=1}^{m} x_i \right) \|$$

$$= \| x - y \|$$

$$\boxed{\Rightarrow L = 1}$$

μ - convexity

$$\nabla^2 f(x) = 1 \quad \Rightarrow \quad \boxed{\mu = 1}$$

x^*

$$\nabla f(x^*) = 0 \quad \Rightarrow \quad 0 = x^* - \frac{1}{m} \sum_{i=1}^{m} x_i$$

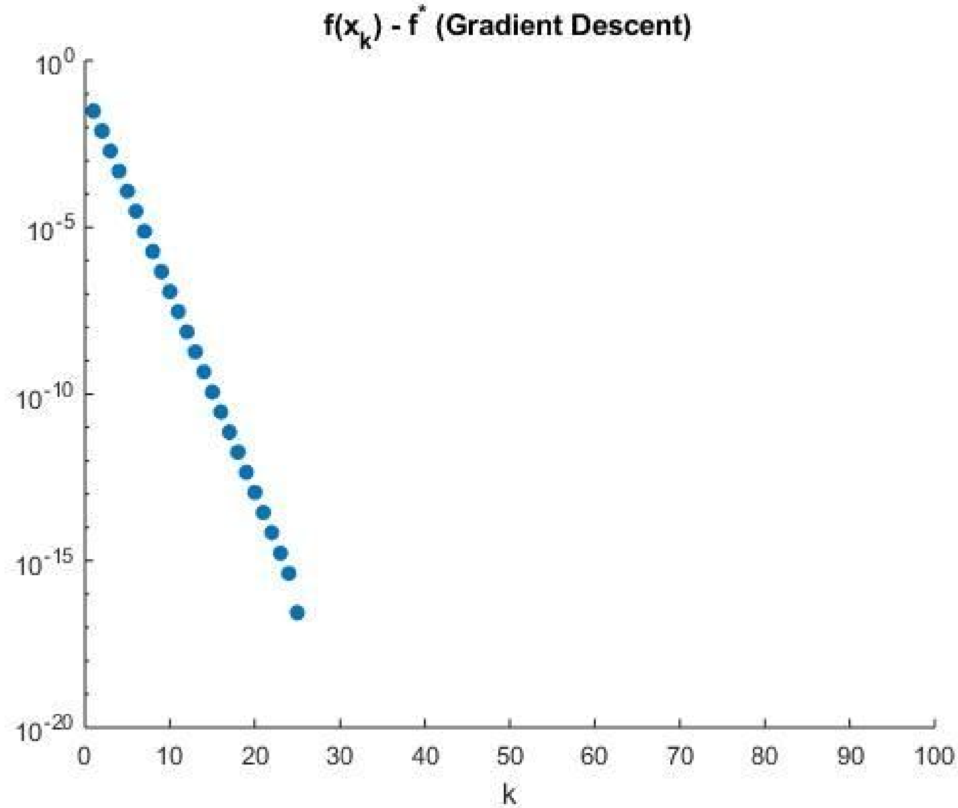
$$\boxed{x^* = \frac{1}{m} \sum_{i=1}^{m} x_i}$$

$f(x^*)$

$$f(x^*) = \frac{1}{2m} \sum_{i=1}^{m} (x^* - x_i)^2$$

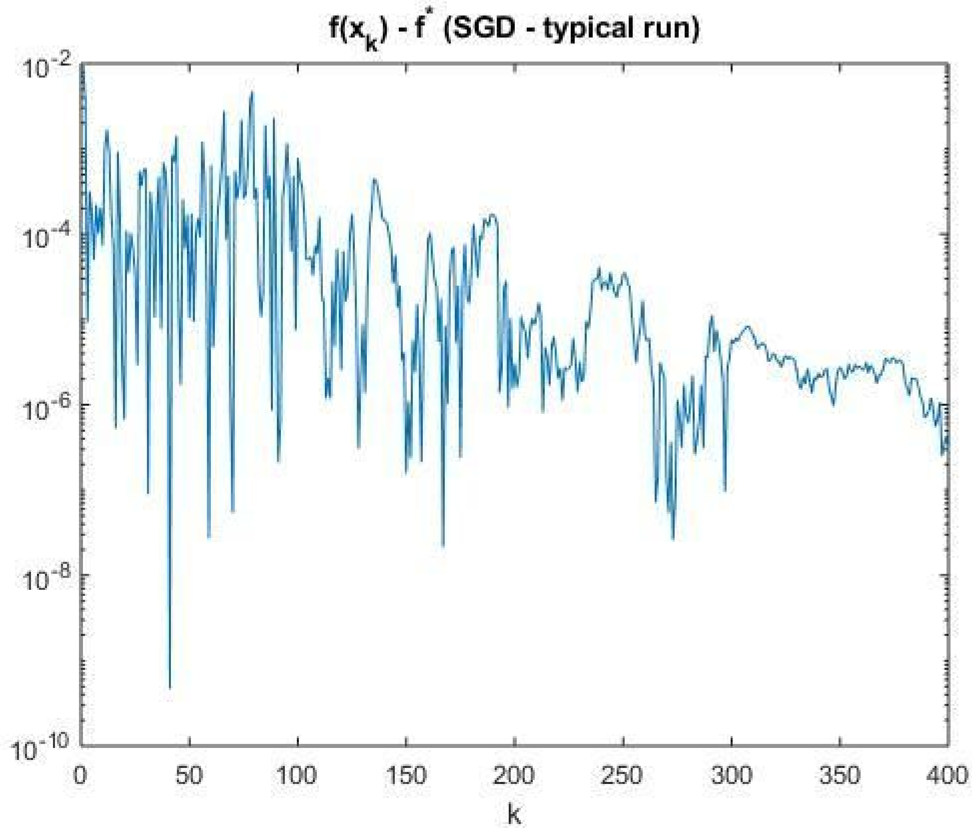
$$\boxed{f(x^*) = \frac{\sigma^2}{2} \quad \text{where} \quad \sigma^2 = Var(x_i) \quad x_i \in x_1 \dots x_m}$$

4B)

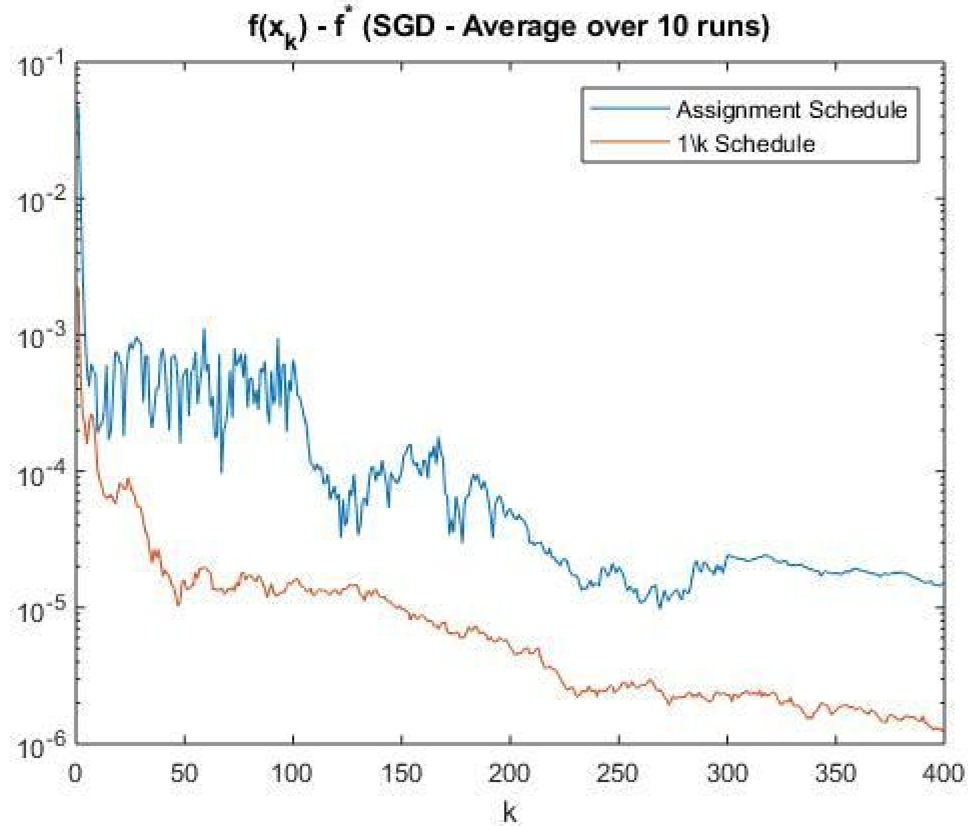


After k=25, MATLAB floating point precision considered $f(x_k) = f^*$. Therefore $\log(f(x_k) - f^*)$ could not be computed

4C)



4D)



I tuned SGD with a schedule of 1/k, where k is the iteration number.

5) $f(x) = \frac{\|Mx - b\|^2}{2}$

$$= \frac{1}{2}(Mx - b)^T(Mx - b)$$

$$= \frac{1}{2}(x^TM^T - b^T)(Mx - b)$$

$$= \frac{1}{2}\left[x^TM^TMx - 2x^TM^Tb + b^Tb\right]$$

$$\nabla f(x) = \frac{1}{2}\left[2M^TMx - 2M^Tb\right]$$

$$= M^TMx - M^Tb$$

$$\nabla^2 f(x) = M^TM$$

M^TM is a Square Symmetric matrix

From class we know that the best μ-convexity and L-Smoothness constants for a quadratic function

$q(x) = \frac{1}{2}x^THx + bx + c$ is $\mu = \lambda_{min}(H) \geq 0$ and

$L = \lambda_{max}(H)$

In this question we have $H = M^TM$

$$\Rightarrow \boxed{\begin{aligned} \mu &= \lambda_{min}(M^TM) \geq 0 \\ L &= \lambda_{max}(M^TM) \end{aligned}}$$