
The McDiarmid Inequality

Recap

In the last scribe, we discussed concentration bounds for infinite hypothesis classes which can be represented by an ϵ -cover, using Hoeffding's Inequality. However, the concentration bound derived was dependent on the dimensionality of the feature space. Thus, the concentration bound derived from Hoeffding's Inequality and ϵ -cover are useless for problems with high dimensionality space such as Kernel Method.

We, therefore need a stronger and more general bound. We therefore study the McDiarmid Inequality, which is a more general result than the Hoeffding's Inequality.

The McDiarmid Inequality

The McDiarmid Inequality provides a more general bound than the Hoeffding's Inequality in the sense that it is independent of the dimensionality of any data point while computing the concentration bound, and thus can be used, even for high dimensional feature spaces.

Before we move on to understand the inequality, we must define what is c -stable function.

Definition 8.1 (c -stable function). Let $X_1, X_2 \dots X_n \in \mathcal{X}^n$ be a set of $n \geq 1$ independent random variables and a function $f : \mathcal{X}^n \rightarrow \mathbb{R}$, then the function f is c -stable if $\exists c \in \mathbb{R}^+$ such that

$$|f(x_1 \dots x_i \dots x_n) - f(x_1 \dots x'_i \dots x_n)| \leq c \quad (1)$$

for all $i \in [1, n]$ and any points $x_1, x_2 \dots x_n, x'_i \in \mathcal{X}$.

We can consider a simple example for a function that is c -stable.

Example 8.1. Suppose we have a loss function \mathcal{X} such that $\mathcal{X} \in [a, b]$ a.s. and a function $f : \mathcal{X}^n \rightarrow \mathbb{R}$ such that

$$f(x_1, x_2 \dots x_n) = \frac{\sum_{i=1}^n x_i}{n}$$

then, we can argue that

$$\begin{aligned}
|f(x_1 \dots x_i \dots x_n) - f(x_1 \dots x'_i \dots x_n)| &= \left| \frac{1}{n} \left[\sum_{j=1}^n x_j - \sum_{\substack{j=1 \\ j \neq i}}^n x_j - x'_i \right] \right| \\
&= \frac{1}{n} \left| \left[\sum_{j=1}^n x_j - \sum_{j=1}^n x_j + x_i - x'_i \right] \right| \\
&= \frac{1}{n} |x_i - x'_i| \\
&\leq \frac{b-a}{n}
\end{aligned}$$

Therefore, f is $\frac{b-a}{n}$ stable.

We can now state the McDiarmid Inequality, which is defined for only c-stable functions.

Theorem 8.1 (McDiarmid Inequality). Let $X_1, X_2 \dots X_n \in \mathcal{X}^n$ be a set of $n \geq 1$ independent random variables and a function $f : \mathcal{X}^n \rightarrow \mathbb{R}$, such that the function f is c-stable. For some sample \mathcal{S} , let $f(\mathcal{S})$ denote $f(X_1, X_2 \dots X_n)$, then the McDiarmid Inequality states that the for all $\epsilon > 0$,

$$\mathbb{P} \left[|f(\mathcal{S}) - \mathbb{E}[f(\mathcal{S})]| > \epsilon \right] \leq 2 \exp \left(\frac{-2\epsilon^2}{nc^2} \right) \quad (2)$$

Proof. Let us first prove the inequality only samples with only a single variable, *i.e.* $n = 1$.

Suppose for a feature space \mathcal{X} , we have a random variable $X \in \mathcal{X}$ and a function $f : \mathcal{X} \rightarrow \mathbb{R}$ such that f is c-stable. Then, for all $x \in \mathcal{X}$, looking at only the right tail (*i.e.* $f(X) - \mathbb{E}[f(X)] > \epsilon$), we can write

$$\begin{aligned}
\mathbb{P} [f(X) - \mathbb{E}[f(X)] > \epsilon] &= \mathbb{P} \left[\exp \left(s \left(f(X) - \mathbb{E}[f(X)] \right) \right) > \exp(s\epsilon) \right] \quad [s > 0] \\
&\leq \frac{\mathbb{E} \left[\exp \left(s \left(f(X) - \mathbb{E}[f(X)] \right) \right) \right]}{\exp(s\epsilon)} \quad [\text{Using Markov's Inequality}]
\end{aligned}$$

Note that $\mathbb{E}[f(X)]$ is always a constant with respect to X . Therefore, we can say

$$\begin{aligned}
\left(\max_X f(X) - \mathbb{E}[f(X)] \right) - \left(\min_X f(X) - \mathbb{E}[f(X)] \right) &= \left(\max_X f(X) \right) - \left(\min_X f(X) \right) \\
&= \max_{X, X'} (f(X) - f(X')) \leq c
\end{aligned}$$

Since f is c-stable, we can say that the difference between the maximum and the minimum values of the quantity $f(X) - \mathbb{E}[f(X)]$ is at most equal to c . Also, since $\mathbb{E}[f(X) - \mathbb{E}[f(X)]] = \mathbb{E}[f(X)] - \mathbb{E}[f(X)] = 0$.

Therefore, we can use Hoeffding's Lemma, discussed in Lecture 5, on the transformed random variable $f(X) - \mathbb{E}[f(X)]$. Hence, we get

$$\mathbb{E} \left[\exp \left(s \left(f(X) - \mathbb{E}[f(X)] \right) \right) \right] \leq \exp \left(\frac{s^2 c^2}{8} \right)$$

Putting this in the original inequality, we get

$$\mathbb{P} [f(X) - \mathbb{E}[f(X)] > \epsilon] \leq \exp \left(\frac{s^2 c^2}{8} - s\epsilon \right)$$

Since we want the closest bound, we minimize the RHS with respect to s . Note that this will be smallest for $s = 4\epsilon/c^2$, therefore

$$\mathbb{P} \left[f(X) - \mathbb{E}[f(X)] > \epsilon \right] \leq \exp \left(-\frac{2\epsilon^2}{c^2} \right)$$

We can follow the same set of steps to prove that $\mathbb{P} \left[f(X) - \mathbb{E}[f(X)] < -\epsilon \right] \leq \exp \left(-2\epsilon^2/c^2 \right)$. Combining these two inequalities, we get

$$\mathbb{P} \left[\left| f(X) - \mathbb{E}[f(X)] \right| > \epsilon \right] \leq 2\exp \left(-\frac{2\epsilon^2}{c^2} \right)$$

This is exactly the term of the McDiarmid Inequality when $n = 1$. Therefore, we can claim that the McDiarmid Inequality holds when $n = 1$. \square

In order to extend the proof to larger training samples, we need what is known as the Martingales. We shall look at this in the next section.

Martingales

In probability theory, a martingale is a sequence of random variables (i.e., a stochastic process) for which, at a particular time in the realized sequence, the expectation of the next value in the sequence is equal to the present observed value, given the knowledge of all prior observed values. (Wikipedia)

We define this more formally below.

Definition 8.2 (Martingale). A sequence of random variables $\{X_i\}_{i=1,2,\dots}$ is said to be a *martingale*, if for all $i = 1, 2, \dots$, the expectation $\mathbb{E}[X_i]$ exists and the conditional expectation

$$\mathbb{E}[X_{i+1} - X_i \mid X_1, X_2 \dots X_i] = 0$$

In order to better understand this, consider the following example of a martingale sequence

Example 8.2 (Random Walk). Suppose a drunkard started his walk from a fixed, known position (say the origin) in 1-D. At any second, he may choose one direction, i.e. left or right at random, and take one step and one step only. If we define the position of the drunkard (a random variable) at any second i as Z_i (let Z_1 be the initial position), then we can say that the sequence $\{Z_i\}_{i=1,2,\dots}$ is a martingale sequence.

It is easy to see why indeed this is a martingale sequence. We can write the expectation

$$\begin{aligned} \mathbb{E}[Z_{i+1} - Z_i \mid Z_1, Z_2 \dots Z_i] &= \mathbb{E}[Z_{i+1} \mid Z_i] - Z_i \\ &= \frac{1}{2}(Z_i + 1) + \frac{1}{2}(Z_i - 1) - Z_i \\ &= 0 \end{aligned}$$

Since the condition for this sequence to be a martingale holds true and therefore, the random walk sequence is a martingale sequence.

Definition 8.3 (Martingale Difference). A sequence of random variables $\{Z_i\}_{i=1,2,\dots}$ is said to be a *martingale difference* with respect to another sequence of random variables $\{X_i\}_{i=1,2,\dots}$ if for all $i = 1, 2, \dots$, Z_i is a function $X_1, X_2 \dots X_i$, $\mathbb{E}[Z_i]$ exists, and

$$\mathbb{E}[Z_{i+1} - Z_i \mid X_1, X_2 \dots X_i] = 0$$

We will derive a concentration bound for a sequence that is a martingale difference with respect to another sequence. Using this, we can prove the McDiarmid Inequality.

Theorem 8.2. Let $\{Z_i\}_{i=1,2,\dots}$ be a martingale difference with respect to another sequence $\{X_i\}_{i=1,2,\dots}$. If for all $i = 1, 2, \dots$

$$a \leq (Z_{i+1} - Z_i \mid X_1 \dots X_i) \leq b$$

then, for all $\epsilon > 0$,

$$\mathbb{P}[|Z_{n+1} - Z_1| > \epsilon] \leq 2 \exp\left(-\frac{2\epsilon^2}{n(b-a)^2}\right) \quad (3)$$

Proof. We borrow the proof technique from the proofs of the Multiplicative Chernoff's Bound and Hoeffding's Bound.

$$\begin{aligned} \mathbb{P}[(Z_n - Z_1) > \epsilon] &= \mathbb{P}[s(Z_n - Z_1) > s\epsilon] \\ &= \mathbb{P}[\exp(s(Z_{n+1} - Z_1)) > \exp(s\epsilon)] \\ &\leq \frac{\mathbb{E}[\exp(s(Z_{n+1} - Z_1))]}{\exp(s\epsilon)} \end{aligned}$$

Let us represent $\{X_j\}_{j=1,2,\dots,i}$ as \bar{X}_i .

Since for all $i = 1, 2, \dots$, the random variable Z_i is a function of \bar{X}_i (from the definition of martingales), we can therefore write the expectation term as below

$$\begin{aligned} \mathbb{E}[\exp(s(Z_{n+1} - Z_1))] &= \mathbb{E}_{\bar{X}_{n+1}} \left[\exp\left(s \left(\sum_{i=1}^n Z_{i+1} - Z_i \right)\right) \right] \\ &\leq \mathbb{E}_{\bar{X}_{n+1}} \left[\prod_{i=1}^n \exp(s(Z_{i+1} - Z_i)) \right] \\ &= \mathbb{E}_{\bar{X}_n} \left[\mathbb{E}_{X_{n+1}} \left[\prod_{i=1}^n \exp(s(Z_{i+1} - Z_i)) \mid \bar{X}_n \right] \right] \end{aligned}$$

Since for all $i = 1, 2, \dots, n$ we can say Z_j is a function of \bar{X}_n , therefore, the product terms for $i = 1 \dots n-1$ will be constant with respect to the expectation, and hence will come out of the inner expectation.

$$\mathbb{E}[\exp(s(Z_{n+1} - Z_1))] \leq \mathbb{E}_{\bar{X}_n} \left[\prod_{i=1}^{n-1} \exp(s(Z_{i+1} - Z_i)) \mathbb{E}_{X_{n+1}} [\exp(s(Z_{n+1} - Z_n)) \mid \bar{X}_n] \right]$$

We can apply Hoeffding's Lemma on $(Z_{n+1} - Z_n)$ as given \bar{X}_n , we know $(Z_n - Z_1) \in [a, b]$ and $\mathbb{E}[Z_{n+1} - Z_n \mid \bar{X}_n] = 0$. Therefore

$$\mathbb{E}[\exp(s(Z_{n+1} - Z_n)) \mid \bar{X}_n] \leq \exp\left(\frac{s^2(b-a)^2}{8}\right)$$

Since, the expectation will always be non-negative, the inequality will hold, and hence, we can write

$$\mathbb{E}[\exp(s(Z_{n+1} - Z_n)) \mid \bar{X}_n] \leq \mathbb{E}_{\bar{X}_n} \left[\prod_{i=1}^{n-1} \exp(s(Z_{i+1} - Z_i)) \mid \bar{X}_n \right] \exp\left(\frac{s^2(b-a)^2}{8}\right)$$

Iterating the same steps, we can therefore reduce the inequality to as written below.

$$\mathbb{E}[\exp(s(Z_n - Z_1))] \leq \exp\left(\frac{ns^2(b-a)^2}{8}\right)$$

Formulating the detailed steps for above is left as an exercise Putting this in the original equation, we get

$$\mathbb{P}[(Z_n - Z_1) > \epsilon] \leq \exp\left(\frac{ns^2(b-a)^2}{8} - s\epsilon\right)$$

Minimizing the upper bound with respect to s , we get $s = 4\epsilon/(b-a)^2$, therefore

$$\mathbb{P}[(Z_n - Z_1) > \epsilon] \leq \exp\left(-\frac{2\epsilon^2}{n(b-a)^2}\right)$$

We can follow the steps to prove that

$$\mathbb{P}[Z_{n+1} - Z_1 < \epsilon] \leq \exp\left(-\frac{2\epsilon^2}{n(b-a)^2}\right)$$

Therefore, combining the two inequalities, we get

$$\mathbb{P}[|Z_{n+1} - Z_1| > \epsilon] \leq 2\exp\left(-\frac{2\epsilon^2}{n(b-a)^2}\right)$$

□

Note. The above theorem is a direct corollary of Azuma's Inequality (Mohri et al., 2012)

Using this theorem, we can finally prove the McDiarmid Inequality, as we will see in the next section

Proof and Corollaries of the McDiarmid Inequality

We can now prove the McDiarmid Inequality using the concept of martingales

Note. We follow the same notation from the previous section, *i.e.* \bar{X}_n represents $X_1, X_2 \dots X_n$

Following from the statement of the McDiarmid Inequality, let $X_1, X_2 \dots X_n \in \mathcal{X}^n$ be a set of independent random variables and $f : \mathcal{X}^n \rightarrow \mathbb{R}$ be a c-stable function Assume another sequence of random variables, \bar{Z}_{n+1} such that for all $i = 0, 1 \dots n-1$

$$Z_i = \mathbb{E}_{X_{i+1} \dots X_n} [f(X_1, X_2 \dots X_n) | \bar{X}_i]$$

and $Z_n = f(X_1, X_2 \dots X_n)$

We therefore want to show

$$\mathbb{E}[|Z_n - Z_0| > \epsilon] \leq 2\exp\left(-\frac{2\epsilon^2}{nc^2}\right)$$

Before showing this, let us consider the following claim

Claim 8.3. The sequence \bar{Z}_n is a martingale difference with respect to \bar{X}_n .

Proof. In order to prove that \bar{Z}_n is a martingale difference with respect to \bar{X}_n , we need to show

$$\mathbb{E}[Z_{i+1} - Z_i | \bar{X}_i] = 0$$

For simplicity, we represent $X_{i+1} \dots X_n$ as \bar{X}_{-i}

$$\begin{aligned}\mathbb{E}[Z_{i+1} - Z_i | \bar{X}_i] &= \mathbb{E}[Z_{i+1} | \bar{X}_i] - \mathbb{E}[Z_i | \bar{X}_i] \\ &= \frac{\mathbb{E}}{\bar{X}_{-i}}[Z_{i+1} | \bar{X}_i] - \frac{\mathbb{E}}{\bar{X}_{-i}}[Z_i | \bar{X}_i] \\ &= \frac{\mathbb{E}}{\bar{X}_{-i}} \left[\frac{\mathbb{E}}{\bar{X}_{-(i+1)}}[f(X_1 \dots X_n) | \bar{X}_{i+1}] | \bar{X}_i \right] - Z_i\end{aligned}$$

Using law of iterated expectation (Rastogi, 2007)

$$\mathbb{E}[A | B] = \mathbb{E}[\mathbb{E}[A | B, C] | B]$$

Therefore,

$$\begin{aligned}\frac{\mathbb{E}}{\bar{X}_{-i}} \left[\frac{\mathbb{E}}{\bar{X}_{-(i+1)}}[f(X_1 \dots X_n) | \bar{X}_{i+1}] | \bar{X}_i \right] &= \frac{\mathbb{E}}{\bar{X}_{-i}}[f(X_1 \dots X_n) | \bar{X}_i] \\ &= Z_i \\ \implies \mathbb{E}[Z_{i+1} - Z_i | \bar{X}_i] &= 0\end{aligned}$$

Hence, our claim is true. \square

Remark 8.1. We can also include X_0 and Z_0 in the sequence where X_0 can be any random variable Since Z_1 is independent of X_0 , $\mathbb{E}[Z_1 - Z_0 | X_0] = \mathbb{E}[Z_1 - Z_0] = 0$ Hence, the sequence $Z_0, Z_1 \dots Z_n$ is a martingale difference with respect to the sequence $X_0, X_1 \dots X_n$.

Exercise 8.1. Prove that since f is c -stable, for all $i = 0, 1 \dots n-1$, $|Z_{i+1} - Z_i| \leq c$.

Since we have shown that the sequence $Z_0 \dots Z_n$ is a martingale difference with respect to the sequence $X_0 \dots X_n$, we can use Theorem 8.2 Also from Exercise 8.1, we know that $|Z_{i+1} - Z_i| \leq c$, we can apply Theorem 8.2 From that, we get

$$\mathbb{P}[|Z_n - Z_0| > \epsilon] \leq 2\exp\left(-\frac{\epsilon^2}{2nc^2}\right)$$

However, note that this is not the same as the expression given for the McDiarmid Inequality We can find a tighter bound by realizing the difference between the maximum and the minimum values of $Z_{i+1} - Z_i$ given \bar{X}_i .

Let $W_i = \sup\{Z_{i+1} - Z_i\}$ given \bar{X}_i and $U_i = \inf\{Z_{i+1} - Z_i\}$

Since we are given \bar{X}_i , therefore Z_i is a constant (as Z_i is a function of \bar{X}_i) Also, since we are computing expectation over $\bar{X}_{-(i+1)}$ in Z_{i+1} , the only variable in question is X_{i+1} Therefore, we can rewrite W_i and U_i as

$$\begin{aligned}W_i &= \sup_x \left\{ \frac{\mathbb{E}}{\bar{X}_{-(i+1)}}[f(S) | X_1, X_2 \dots X_i, x] \right\} - Z_i \\ U_i &= \inf_x \left\{ \frac{\mathbb{E}}{\bar{X}_{-(i+1)}}[f(S) | X_1, X_2 \dots X_i, x] \right\} - Z_i\end{aligned}$$

Hence, the difference can be written as

$$\begin{aligned} \implies W_i - U_i &= \sup_x \left\{ \mathbb{E}_{\bar{X}_{-(i+1)}} [f(S) | X_1, X_2 \dots X_i, x] \right\} - \inf_x \left\{ \mathbb{E}_{\bar{X}_{-(i+1)}} [f(S) | X_1, X_2 \dots X_i, x] \right\} \\ &= \sup_{x, x'} \left\{ \mathbb{E}_{\bar{X}_{-(i+1)}} [f(S) | X_1, X_2 \dots X_i, x] - \mathbb{E}_{\bar{X}_{-(i+1)}} [f(S) | X_1, X_2 \dots X_i, x'] \right\} \end{aligned}$$

This, similar to Exercise 8.1 can be proved to be smaller than c Also,

$$U_i \leq Z_{i+1} - Z_i \leq W_i$$

Since U_i is a constant given \bar{X}_i as it does not depend on \bar{X}_{-i} , and $W_i - U_i \leq c$, we have

$$U_i \leq Z_{i+1} - Z_i \leq U_i + c$$

Applying Theorem 8.2, we get the actual term for the McDiarmid Inequality, *i.e.*

$$\begin{aligned} \mathbb{P}[|Z_n - Z_0|] &\leq \exp\left(-\frac{2n\epsilon^2}{c^2}\right) \\ \implies \mathbb{P}\left[\left|f(X_1, X_2 \dots X_n) - \mathbb{E}[f(X_1, X_2 \dots X_n)]\right|\right] &\leq \exp\left(-\frac{2n\epsilon^2}{c^2}\right) \end{aligned}$$

This completes the proof of the McDiarmid Inequality Below, we see some interesting applications / corollaries of the McDiarmid Inequality

Remark 8.2. We can use the McDiarmid Inequality for functions which are c -stable with high probability, but not almost surely.

Corollary 8.2 (Hoeffding's Inequality). Consider the function discussed in Example 8.1, *i.e.* $f(X_1 \dots X_n) = \frac{1}{n} \sum_{i=1}^n X_i$ Since, f is $\frac{b-a}{n}$ -stable, we can apply the the McDiarmid Inequality to f , using which we directly get the Hoeffding's Inequality The details of the intermediate steps are left as an exercise.

Corollary 8.3. If for some loss function $l : \mathcal{X} \times \mathcal{Y}$, we have $\forall (x, y) \in \mathcal{X} \times \mathcal{Y}$ and $\forall f, l(f(x), y) \in [a, b]$, then for some sample $\mathcal{S} = \{(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)\}$, we have

$$\mathbb{P}\left[\left|g(\mathcal{S}) - \mathbb{E}[g(\mathcal{S})]\right| > \epsilon\right] \leq \exp\left(-\frac{2n\epsilon}{(b-a)^2}\right) \quad (4)$$

where $g(\mathcal{S})$ is defined as follows

$$g(\mathcal{S}) = \sup_{f \in \mathcal{F}} \{ \text{er}_{\mathcal{S}}^l[f] - \text{er}_{\mathcal{D}}^l[f] \}$$

Proof. Suppose we have a loss function $l : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$, such that for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$ and $\forall f \in \mathcal{F}$, we have

$$a \leq l(f(x), y) \leq b$$

Suppose we have a training sample $S = \{(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)\}$

We want to show that g is c -stable, which would allow us to use the McDiarmid Inequality for this function.

Define $g_f(S) = \text{er}_{\mathcal{S}}^l[f] - \text{er}_{\mathcal{D}}^l[f]$ and $S_i = S \setminus \{(x_i, y_i)\} \cup \{(x'_i, y_i)\}$

Remark 8.3. $g(\mathcal{S})$, information theoretical wise, is well defined, even though we cannot calculate the value of $g(\mathcal{S})$ (since we cannot compute $\text{er}_{\mathcal{D}}^l[f]$) algorithmically.

Since $l(f(x), y) \in [a, b]$, we have for all $i \in [n]$

$$\begin{aligned} g_f(S) - g_f(S_i) &= \frac{1}{n} \left(\sum_{j=1}^n l(f(x_j), y_j) - \left(\sum_{\substack{j=1 \\ j \neq i}}^n l(f(x_j), y_j) + l(f(x'_i), y_i) \right) \right) \\ &= \frac{1}{n} (l(f(x_i), y_i) - l(f(x'_i), y_i)) \end{aligned}$$

Since $l(f(x), y) \in [a, b]$

$$g_f(S) - g_f(S_i) \leq \frac{b-a}{n} \quad (5)$$

Therefore, we can write, for all $i \in [n]$

$$g(S) - g(S_i) = \sup_{f \in \mathcal{F}} g_f(S) - \sup_{f \in \mathcal{F}} g_f(S_i)$$

Without loss of generality, say $g(S)$ is maximum for f_1 and $g(S_i)$ is maximum for f_2

$$\begin{aligned} g(S) - g(S_i) &= g_{f_1}(S) - g_{f_2}(S_i) \\ &\leq g_{f_1}(S) - g_{f_1}(S_i) + \frac{b-a}{n} \quad [\text{Using Equation 5}] \\ &= \frac{b-a}{n} \end{aligned}$$

Therefore, g is $\frac{b-a}{n}$ -stable This suggests that we can use the McDiarmid Inequality for g Hence, we get

$$\mathbb{P} \left[g(S) - \mathbb{E}_S[g(S)] > \epsilon \right] \leq \exp \left(-\frac{2n\epsilon^2}{(b-a)^2} \right)$$

Using similar steps, we can prove the left tail for this as well. \square

The above corollary, at first glance, seems like the condition for uniform convergence. However, one needs to realize that the expectation of the function $g(S)$ is not necessarily zero. This is because the function $g(S)$ is over a supremum, which makes life a bit harder for us.

In order to really find a bound on $g(S)$, we need to, in turn, bound the value of its expectation, however this is not as trivial. We will, in fact, see this in the next lecture.

Nonetheless, one see that the McDiarmid Inequality is really useful for finding bounds for c -stable functions, even for infinite hypothesis classes, since the bound is independent of the size of the hypothesis class, unlike concentration inequalities discussed until now.

In the next lecture, we will revise the bound for the regression case using the McDiarmid Inequality and see that for certain hypothesis classes, it is possible to get a bound on the number of samples required for a model ϵ -close to the real data distribution, by bounding the value of $\mathbb{E}[g(S)]$.

References

- Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. The MIT Press, 2012.
- Ashish Rastogi. *McDiarmid Inequality*. 2007.
- Wikipedia. Martingale (Probability Theory). [https://en.wikipedia.org/wiki/Martingale_\(probability_theory\)](https://en.wikipedia.org/wiki/Martingale_(probability_theory)).