Fundamentals of Statistical Learning; Prof. Dr. Celisse Alain; Homework 2, Exercise 2

Anton Zaitsev | 0230981826 | anton.zaitsev.001@student.uni.lu | University of Luxembourg

May 2, 2024

Exercise 2

1.

First, let us study how $K_h(X_1 - x_0)$ is bounded:

$$|K_h(X_1 - x_0)| = \left|\frac{1}{h}K\left(\frac{X_i - x}{h}\right)\right| \leq \left|\frac{1}{h}||K||_\infty\right| = \frac{1}{h}||K||_\infty$$

Second, let us study how the expectation of $K_h(X_1 - x_0)$ is bounded:

$$|\mathbb{E}\left[K_h(X_1 - x_0)\right]| \overset{\text{Jensen}}{\leq} \mathbb{E}\left[|K_h(X_1 - x_0)|\right] \leq \mathbb{E}\left[\frac{||K||_\infty}{h}\right] = \frac{1}{h}||K||_\infty$$

We get that:

$$0 < |K_h(X_1 - x_0) - \mathbb{E}\left[K_h(X_1 - x_0)\right]| \leq \frac{2}{h}||K||_\infty$$

Now, we can express the bound of $\mathbb{E}\left[|\zeta_1|^k\right]$:

$$\begin{aligned}
\mathbb{E}\left[|\zeta_1|^k\right] &= \mathbb{E}\left[|K_h(X_1 - x_0) - \mathbb{E}\left[K_h(X_1 - x_0)\right]|^k\right] \\
&\leq \left(\frac{2}{h}||K||_\infty\right)^k \\
&= \left(\frac{2}{h}||K||_\infty\right)^{k-2} \mathbb{E}\left[|K_h(X_1 - x_0) - \mathbb{E}\left[K_h(X_1 - x_0)\right]|^2\right] \\
&= \left(\frac{2}{h}||K||_\infty\right)^{k-2} Var(K_h(X_1 - x_0)) \\
&\leq \frac{k!}{2}\left(\frac{2}{h}||K||_\infty\right)^{k-2} Var(K_h(X_1 - x_0)) \quad \text{for } k \geq 2
\end{aligned}$$

2.

In **1.** we proved that ζ_1 satisfies $BC(v, c)$. Notice that since X_i's are independent and identically distributed real-valued random variables, $\zeta_1, \ldots, \zeta_n$ are also n independent and identically distributed random variables by the definition of ζ_i. Thus, we can use the Bernstein's inequality and get that for every $t > 0$:

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}\zeta_i > t\right] \vee \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}\zeta_i < -t\right] \leq e^{\frac{-nt^2}{2(v+ct)}}$$

$$\iff$$

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n}\zeta_i\right| > t\right] \leq 2e^{\frac{-nt^2}{2(v+ct)}}$$

$$\iff$$

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n}\left(K_h(X_i - x_0) - \mathbb{E}[K_h(X_i - x_0)]\right)\right| > t\right] \leq 2e^{\frac{-nt^2}{2(v+ct)}}$$

$$\iff$$

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n}K_h(X_i-x_0)-\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}K_h(X_i-x_0)\right]\right|>t\right]\leq 2e^{\frac{-nt^2}{2(v+ct)}}$$

$$\iff$$

$$\mathbb{P}\left[\left|\hat{f}_h(x_0)-\mathbb{E}\left[\hat{f}_h(x_0)\right]\right|>t\right]\leq 2e^{\frac{-nt^2}{2(v+ct)}}$$

3.

We need to find y s.t. $x=\frac{ny^2}{2(v+cy)}$. Then:

$$x=\frac{ny^2}{2(v+cy)}$$

$$\iff$$

$$y^2-\frac{2cx}{n}y-\frac{2xv}{n}=0 => y=\frac{1}{2}\left(\frac{2cx}{n}\pm\sqrt{\frac{4c^2x^2}{n^2}+4\frac{2xv}{n}}\right)=\frac{cx}{n}\pm\sqrt{\frac{c^2x^2}{n^2}+\frac{2xv}{n}}$$

We take $y=\frac{cx}{n}+\sqrt{\frac{c^2x^2}{n^2}+\frac{2xv}{n}}$ so that $y>0$. We then get:

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}X_i>y\right]\leq e^{\frac{-ny^2}{2(v+cy)}}$$

with $y=\frac{cx}{n}+\sqrt{\frac{c^2x^2}{n^2}+\frac{2xv}{n}}$:

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}X_i>\frac{cx}{n}+\sqrt{\frac{c^2x^2}{n^2}+\frac{2xv}{n}}\right]\leq e^{-x}\quad v,c,x>0$$

Solving 3. using the solution from the exercise sheet

Note that with this solution we only get an implication and not an equivalence as required in the exercise sheet.

Let $y=\sqrt{\frac{2vx}{n}}+\frac{cx}{n}, v,c,x>0$. Then:

$$\begin{aligned}e^{\frac{-ny^2}{2(v+cy)}}&=e^{\frac{-n\left(\frac{2vx}{n}+\frac{2cx}{n}\sqrt{\frac{2vx}{n}}+\frac{c^2x^2}{n^2}\right)}{2v+2c\sqrt{\frac{2vx}{n}}+\frac{2c^2x}{n}}}\\&=e^{\frac{-x\left(2v+2c\sqrt{\frac{2vx}{n}}+\frac{c^2x}{n}\right)}{2v+2c\sqrt{\frac{2vx}{n}}+\frac{2c^2x}{n}}}\\&\leq e^{-x}\quad\text{hence implication, not equivalence}\end{aligned}$$

Notice that $2v+2c\sqrt{\frac{2vx}{n}}+\frac{c^2x}{n}<=2v+2c\sqrt{\frac{2vx}{n}}+\frac{2c^2x}{n}$ since $c,x>0$. Thus:

$$\begin{aligned}e^{\frac{-ny^2}{2(v+cy)}}&=e^{\frac{-x(2v+2c\sqrt{\frac{2vx}{n}}+\frac{c^2x}{n})}{2v+2c\sqrt{\frac{2vx}{n}}+\frac{2c^2x}{n}}}\\&\leq e^{-x}\end{aligned}$$

Thus, for $y>0$:

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} X_i > y\right] \leq e^{\frac{-ny^2}{2(v+cy)}}$$

with $y = \sqrt{\frac{2vx}{n}} + \frac{cx}{n}, v, c, x > 0$:

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} X_i > \sqrt{\frac{2vx}{n}} + \frac{cx}{n}\right] \leq e^{-x}$$

4.

By using the Bernstein's inequality on ζ_i - iid real-valued random variables, we get

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} \zeta_i > t\right] \leq e^{\frac{-nt^2}{2(v+ct)}} \iff \mathbb{P}\left[\hat{f}_h(x_0) - \mathbb{E}\left[\hat{f}_h(x_0)\right] > t\right] \leq e^{\frac{-nt^2}{2(v+ct)}} \quad \forall x > 0$$

With the above inequality equal to (from **3.**)

$$\mathbb{P}\left[\hat{f}_h(x_0) - \mathbb{E}\left[\hat{f}_h(x_0)\right] > \frac{cx}{n} + \sqrt{\frac{c^2x^2}{n^2} + \frac{2xv}{n}}\right] \leq e^{-x},$$

for $x > 0$.

Let $g(x) = x^2$. Note that to apply g to both sides of the inequality in the probability term, we need to be sure that both of these terms are positive. Then

$$\begin{aligned}\mathbb{P}\left[\left|\hat{f}_h(x_0) - \mathbb{E}\left[\hat{f}_h(x_0)\right]\right| > \frac{cx}{n} + \sqrt{\frac{c^2x^2}{n^2} + \frac{2xv}{n}}\right] &= \mathbb{P}\left[\left(\hat{f}_h(x_0) - \mathbb{E}\left[\hat{f}_h(x_0)\right]\right)^2 > \left(\frac{cx}{n} + \sqrt{\frac{c^2x^2}{n^2} + \frac{2xv}{n}}\right)^2\right] \\ &\leq 2e^{-x} \quad \forall x > 0\end{aligned}$$

Using the fact that for all $a, b \in \mathbb{R}, (a+b)^2 \leq 2(a^2 + b^2)$ we achieve:

$$\mathbb{P}\left[\left(\hat{f}_h(x_0) - \mathbb{E}\left[\hat{f}_h(x_0)\right]\right)^2 > 4\left(\left(\frac{cx}{n}\right)^2 + \frac{xv}{n}\right)\right] \leq 2e^{-x} \quad \forall x > 0$$

Note that the result differs from the one in the exercise sheet. We would achieve the same result as in the exercise sheet if we used $y = \sqrt{\frac{2vx}{n}} + \frac{cx}{n}$ and assumed that $\hat{f}_h(x_0) - \mathbb{E}\left[\hat{f}_h(x_0)\right]$ is positive and followed the same steps as above.

5.

We know that ζ_i are independent. From **1.** we know that ζ_i are bounded by $\frac{2}{h}||K||_\infty$. Notice also that ζ_i are centered, since $\mathbb{E}\left[\zeta_i\right] = \mathbb{E}\left[K_h(X_i - x_0) - \mathbb{E}\left[K_h(X_i - x_0)\right]\right] = \mathbb{E}\left[K_h(X_i - x_0)\right] - \mathbb{E}\left[K_h(X_i - x_0)\right] = 0$. Thus, we have all the requirements to use the Hoeffing's inequality for ζ_i. Recall that Hoeffing's inequality requires that the random variables under consideration are independent, centered and bounded. Then, for all $t > 0$:

$$
\begin{aligned}
\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n}\zeta_i\right| > t\right] &= \mathbb{P}\left[\left|\hat{f}_h(x_0) - \mathbb{E}\left[\hat{f}_h(x_0)\right]\right| > t\right] \\
&\leq 2e^{\frac{-2n^2t^2}{\sum_{i=1}^{n}(b_i-a_i)^2}} \\
&= 2e^{\frac{-2nh^2t^2}{4||K||_\infty^2}} \quad \text{with } a = 0, b = \frac{2}{h}||K||_\infty \\
&= 2e^{\frac{-nt^2}{2c^2}} \quad \text{with } c = \left(\frac{2}{h}||K||_\infty\right)^2
\end{aligned}
$$

6.

We apply Hoeffding's inequality to ζ_i:

$$
\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n}\zeta_i\right| > t\right] = \mathbb{P}\left[\left|\hat{f}_h - \mathbb{E}\left[\hat{f}_h\right]\right| > t\right] \leq 2e^{\frac{-nt^2}{2c^2}} \quad \text{with } c = \left(\frac{2||K||_\infty}{h}\right)^2
$$

Using Bernstein's inequality, we achieved the following bound:

$$
\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n}\zeta_i\right| > t\right] = \mathbb{P}\left[\left|\hat{f}_h - \mathbb{E}\left[\hat{f}_h\right]\right| > t\right] \leq 2e^{\frac{-nt^2}{2(v+ct)}}
$$

where $c = \frac{2||K||_\infty}{h}$ (same as before in Hoeffding's inequality) and $v = Var(K_h(X_1 - x_0))$.

We compare c^2 and $v + ct$.

We can notice that in comparison to the Bernstein's inequality, Hoeffding's inequality considers the highest variance possible, which is c^2. For $v << c^2$, Bernstein's inequality provides a tighter upper bound. On the contrary, when v is close to c^2 or equal to c^2, Hoeffding's inequality actually provides tighter bound. The advantage is limited, since the maximum variance of $K_h(X_1 - x_0)$ is c^2 (bounded).