---

**Exercise sheet 9**

---

**Exercise 1** Show that if $X_0, X_1, \dots, X_n \in \mathbb{R}^d$ are independent and identically distributed with a continuous density function $f$, then for all $u > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(n^{1/d}\|X_{(1)}(X_0) - X_0\| > u | X_0\right) = e^{-V_d f(X_0) u^d} \quad a.s,$$

where $V_d > 0$ is the volume of the unit ball in $\mathbb{R}^d$ ($V_1 = 2$).

**Exercise 2** This question shows how to obtain rates of convergence for the $k$-nearest neighbours classifier under smoothness conditions.

Let $(X_1, Y_1), \dots, (X_n, Y_n) \in \mathbb{R}^d \times \{0,1\}$ be independent and identically distributed and let $\eta(x) = \mathbb{P}(Y_1 = 1 | X_1 = x)$. Let $(X, Y)$ be independent of $\{(X_i, Y_i)\}_{i=1}^n$ and have the same distribution as $(X_1, Y_1)$. As in the lectures define $\tilde{\eta}_n(x) = \mathbb{E}[\hat{\eta}_n(x) | X_1, \dots, X_n]$ for $x \in \mathbb{R}^d$. Suppose that $|\eta(x) - \eta(y)| \le M\|x - y\|$ for all $x, y \in \mathbb{R}^d$. Show that for all $\delta > 0$,

$$\mathbb{E}[(\tilde{\eta}_n(X) - \eta(X))^2] \le 4dM^2 \mathbb{P}(\|X_{(k)}(X) - X\| > \delta) + M^2\delta^2,$$

where the expectation and the probability are taken with respect to all the random variables $(X, Y), (X_1, Y_1), \dots, (X_n, Y_n)$.

Now suppose that $\mathbb{P}(X_1 \in [-1,1]^d) = 1$ and that $X_1$ has a density $f$ that is bounded below on $[-1,1]^d$. For $x \in [-1,1]^d$ and $t \ge 0$ define the function $F_x(t) = \mathbb{P}(\|X_1 - x\| \le t)$. You may use without proof that under these conditions, there exists $c \in (0,1]$, independent of $x$, such that $F_x^{-1}(s) \le (s/c)^{1/d}$ for all $s \in [0, c]$.

Show that if $\delta = (2k/[nc])^{1/d} \le 1$ then for all $x \in [-1,1]^d$,

$$\mathbb{P}(\|X_{(k)}(x) - x\| > \delta) \le \tfrac{1}{k}.$$

*Hint: use exercise 5 from week 8.*

Show that there exist finite positive $C_1, C_2, \epsilon$, independent of $k$ and $n$, such that if $k/n \le \epsilon$ then

$$\mathbb{E}[(\tilde{\eta}_n(X) - \eta(X))^2] \le \tfrac{C_1}{k} + C_2 \left(\tfrac{k}{n}\right)^{2/d}.$$

Deduce that for an appropriate choice of $k$,

$$\mathbb{E}\hat{g}_n(X) - R(g^*) \le Cn^{-1/(d+2)}$$

for some finite constant $C$ that does not depend on $n$.

**Bonus.** Assuming that $X$ takes values in $[-1,1]^d$ and has density bounded below there, show that there exists $c \in (0,1]$ such that for all $x \in [-1,1]^d$, the function $F_x(t) = \mathbb{P}(\|X_1 - x\| \le t)$ satisfies

$$F_x(t) \ge ct^d$$

for all $t \le 1$. Deduce an upper bound on its inverse $F_x^{-1}(s)$ for $s \le c$ and all $x \in [-1,1]^d$.

**Exercise 3** Let $Y$ be a nonnegative random variable and $y > 0$. Show that

$$\inf_{t\ge 0} e^{-ty}\mathbb{E}e^{tY} \ge \inf_{k\in\mathbb{N}\cup\{0\}} y^{-k}\mathbb{E}Y^k.$$

In other words, the Chernoff bound can be improved if instead of $e^{tY}$ we consider the moments $Y^k$. When does the inequality hold as equality?

**Exercise 4** Let $Y$ be a random variable with mean zero and $a \leq Y \leq b$. Use convexity to show that for every $t \in \mathbb{R}$, we have

$$\log \mathbb{E}(e^{tY}) \leq -\alpha u + \log(\beta + \alpha e^u),$$

where $u = t(b-a)$ and $\alpha = 1 - \beta = -a/(b-a)$. Using a second-order Taylor expansion about the origin, deduce that $\log \mathbb{E}(e^{tY}) \leq t^2(b-a)^2/8$.

**Exercise 5** Suppose that $\overline{X}_n$ satisfies

$$\mathbb{P}(\overline{X}_n > x) \leq \inf_{t \in [0,1/M)} \exp\left(n \tfrac{t^2\sigma^2}{2(1-tM)}\right) \exp(-tnx)$$

for all $x > 0$.

$$\mathbb{P}\left(\overline{X}_n \geq \tfrac{\sigma\sqrt{2}}{\sqrt{n}} \sqrt{\log \tfrac{1}{\delta}} + \tfrac{M}{n} \log \tfrac{1}{\delta}\right) \leq \delta, \qquad \delta \in (0,1].$$

*Hint: optimise the bound over* $s = 1 - Mt \in (0,1]$