Winter term 2020/21
L. Rendsburg / U. von Luxburg

Assignment 10

Mathematics for Machine Learning

Submission due on **01.02.21, 8:00**

Justify all your claims.

Exercise 1 (**Multivariate distributions**, 3+2 points)**.**

a) Consider the joint density

$$f(x,y) = \frac{1}{Z} \exp\left(-2x^2 - y^2 - x^2y^2\right), \quad x, y \in \mathbb{R},$$

of two real-valued random variables X and Y, where $Z = \int_{\mathbb{R}^2} \exp\left(-2x^2 - y^2 - x^2y^2\right) d(x,y)$ is the normalizing constant. Compute the marginal densities $f_X(x)$ and $f_Y(y)$, as well as the conditional densities $f_{X|Y=y}(x)$ and $f_{Y|X=x}(y)$. What is the name of the distributions given by the conditional densities?
Hint: Use the formula $\int_{\mathbb{R}} \exp(-a(x+b)^2)dx = \sqrt{\frac{\pi}{a}}$ for $a > 0, b \in \mathbb{R}$.

b) Consider a positive joint density $f(x,y)$ of two real-valued random variables X and Y. Prove the continuous versions of Bayes' formula and the law of total probability for all $x, y \in \mathbb{R}$:

$$f_{Y|X=x}(y) = \frac{f_{X|Y=y}(x) f_Y(y)}{f_X(x)} \qquad \text{(Bayes' formula)}$$

and

$$f_Y(y) = \int_{\mathbb{R}} f_{Y|X=x}(y) f_X(x) dx \qquad \text{(Law of total probability)}$$

Exercise 2 (**Concentration inequalities**, 2+3+3+2 points)**.** Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be i.i.d. random variables taking values in $\mathbb{R}^d \times \mathbb{R}$. For a function $g\colon \mathbb{R}^d \times \mathbb{R} \to [0,1]$, define

$$R(g) := \mathbb{E}_{(X,Y)}\left[g(X,Y)\right] \quad \text{and} \quad R_n(g) = \frac{1}{n}\sum_{i=1}^{n} g(X_i, Y_i)\,.$$

Let $\mathcal{G} = \{g_1, \dots, g_m\}$ be a finite set of $m \in \mathbb{N}$ such functions. Define

$$g_n := \operatorname*{arg\,min}_{g \in \mathcal{G}} R_n(g) \quad \text{and} \quad g^* := \operatorname*{arg\,min}_{g \in \mathcal{G}} R(g)\,.$$

a) Let $Z_1, \dots, Z_n$ be i.i.d random variables taking values in $[0,1]$ with $\mu = \mathbb{E}[Z_1]$. Use Hoeffding's inequality to prove that for any $\varepsilon > 0$, it holds

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} Z_i - \mu\right| \geq \varepsilon\right) \leq 2\exp(-2n\varepsilon^2)\,.$$

b) Use part a) on $Z_i := g(X_i, Y_i)$ to prove that for any $\varepsilon > 0$, it holds

$$P\left(\sup_{g \in \mathcal{G}} |R_n(g) - R(g)| \geq \varepsilon\right) \leq 2m\exp(-2n\varepsilon^2)$$

c) Prove that for any $\varepsilon > 0$, it holds

$$P\left(|R(g_n) - R(g^*)| \geq \varepsilon\right) \leq P\left(\sup_{g \in \mathcal{G}} |R_n(g) - R(g)| \geq \frac{\varepsilon}{2}\right)$$

Hint: prove the implication $\left(|R(g_n) - R(g^*)| \geq \varepsilon\right) \Rightarrow \left(\sup_{g \in \mathcal{G}} |R_n(g) - R(g)| \geq \frac{\varepsilon}{2}\right)$ by using the decomposition

$$R(g_n) - R(g^*) = [R(g_n) - R_n(g_n)] + [R_n(g_n) - R_n(g^*)] + [R_n(g^*) - R(g^*)] \ .$$

d) Combine the inequalities from part b) and c) to prove that $R(g_n) \to R(g^*)$ almost surely as $n \to \infty$.
Hint: Use the Borel-Cantelli lemma and the following characterization (compare Assignment 9, Exercise 2c)

$$X_n \xrightarrow[n\to\infty]{} X \text{ almost surely} \quad \Leftrightarrow \quad \forall \varepsilon > 0: \ P(\{|X_n - X| \geq \varepsilon\} \text{ i.o.}) = 0\,.$$

Exercise 3 (**Variance bounds**, 3+2 points)**.** The Efron-Stein inequality is stated as follows. Let $X_1, \ldots, X_n$ be independent random variables taking values in $\mathbb{R}$ and let $Z = f(X_1, \ldots, X_n)$ be a square-integrable function, that is, $\mathbb{E}\left[Z^2\right] < \infty$. Moreover, let $X_1', \ldots, X_n'$ be independent copies of $X_1, \ldots, X_n$, that is, they are jointly independent and X_i' has the same distribution as X_i for every $i \in \{1, \ldots, n\}$. For every i, define Z_i' as the random variable obtained by replacing X_i with X_i', that is,

$$Z_i' = f(X_1, \ldots, X_{i-1}, X_i', X_{i+1}, \ldots, X_n)\,.$$

Then it holds

$$\mathrm{Var}[Z] \leq \sum_{i=1}^{n} \mathbb{E}[(Z - Z_i')_+^2]\,, \quad \text{where } x_+ := \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}\,.$$

Now consider the following scenario. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric real matrix whose entries $A_{i,j} = X_{i,j}$ for $1 \leq i \leq j \leq n$ are independent random variables taking values in $[-1, 1]$. Let $Z = Z(A)$ denote the largest eigenvalue of A.

a) For $1 \leq i \leq j \leq n$, let $X_{i,j}'$ be an independent copy of $X_{i,j}$. Consider the symmetric matrix $A_{i,j}'$ obtained by replacing $X_{i,j}$ in A with $X_{i,j}'$, and let $Z_{i,j}'$ denote the corresponding largest eigenvalue. Let $v = (v_1, \ldots, v_n)^T \in \mathbb{R}^n$ denote an eigenvector of A corresponding to the largest eigenvalue Z with $\|v\| = 1$. Prove that

$$(Z - Z_{i,j}')_+ \leq 4\,|v_i v_j|\ .$$

Hint: Use the fact that the largest eigenvalue Z satisfies

$$Z = v^T A v = \sup_{u \in \mathbb{R}^n\,:\,\|u\|=1} u^T A u \tag{$*$}$$

b) Use part a) to prove that $\mathrm{Var}[Z] \leq 16$.
Hint: You may assume that $\mathbb{E}\left[Z^2\right] < \infty$ without a proof.