Winter term 2024/25
U. von Luxburg
E. Günther/ K. Frohnapfel

Assignment 12

Mathematics for Machine Learning

Submission due Friday **24.01.25, 23:59** via Ilias

Justify all your claims.

Exercise 1 (**Multivariate distributions with densities**, 3+2 points)**.** Let X, Y be two real-valued random variables.

a) Consider the joint density

$$f_{X,Y}(x,y) = \frac{1}{c}\exp\left(-2x^2 - y^2 - x^2y^2\right), \quad x, y \in \mathbb{R},$$

of X and Y, where $c = \int_{\mathbb{R}^2} \exp\left(-2x^2 - y^2 - x^2y^2\right) \mathrm{d}(x,y)$ 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 are the names and parameters of the distributions given by the conditional densities?
Hint: You can use without proof that $\int_{\mathbb{R}} \exp(-a(x+b)^2)\mathrm{d}x = \sqrt{\frac{\pi}{a}}$ *for* $a > 0, b \in \mathbb{R}$.

b) Consider a positive joint density $f_{X,Y}(x,y)$ and 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

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

Exercise 2 (**Conditional Expectation**, 2+2+2 points)**.**
Let $(\Omega, \mathcal{A}, P)$ be a probability space and $X, Y, Z \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$. The conditional covariance of X, Y given Z is defined as

$$\mathrm{Cov}(X, Y \mid Z) := \mathrm{E}(XY \mid Z) - \mathrm{E}(X \mid Z)\mathrm{E}(Y \mid Z).$$

The conditional variance of X given Z is defined as $\mathrm{Var}(X \mid Z) := \mathrm{Cov}(X, X \mid Z)$.

a) The *law of total expectation* states

$$\mathrm{E}(X) = \mathrm{E}(\mathrm{E}(X \mid Y))$$

for two random variables X, Y. Prove this law for two discrete random variables.

b) Prove the *law of total covariance*, which states

$$\mathrm{Cov}(X, Y) = \mathrm{E}(\mathrm{Cov}(X, Y \mid Z)) + \mathrm{Cov}(E(X \mid Z), E(Y \mid Z))$$

and deduce that

$$\mathrm{Var}(X) = \mathrm{E}(\mathrm{Var}(X \mid Z)) + \mathrm{Var}(\mathrm{E}(X \mid Z)),$$

which is known as the *law of total variance.* You may use the law of total expectation, although we only proved it for discrete random variables.

c) What is the interpretation of the formula in b) and the intuitive meaning of the two summands? You can choose to explain either the covariance formula or the variance formula. Find real-life examples for X, Y, Z, where $\mathrm{Cov}(X, Y)$ is non-zero but $\mathrm{E}(\mathrm{Cov}(X, Y \mid Z))$ vanishes. You do not need to make this mathematically rigorous, just explain your example and why the terms behave as they do.

Exercise 3 (**Unbiased estimators are not always useful**, 2+1+2 points)**.**
Let $X \sim \mathrm{Pois}(-0.5 \log \theta)$ be a Poisson-distributed random variable with $\theta \in \Theta = (0, 1)$. We now want to estimate θ based on one sample of X.

a) Consider the estimator $U = (-1)^X$. Prove that U is the only unbiased estimator for θ.
Hint: Use the equality $\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}$. *Additionally, use the fact that*

$$\left(\sum_{k=0}^{\infty} a_k \frac{x^k}{k!} = \sum_{k=0}^{\infty} b_k \frac{x^k}{k!} \quad \forall x \in (0, \infty) \right) \quad \Rightarrow \quad a_k = b_k \quad \forall k \in \mathbb{N}_0 \,.$$

b) What is the $\mathrm{MSE}(U, \theta)$?

c) Now consider another estimator $V = \mathbb{1}_{2\mathbb{N}_0}(X)$. Is V unbiased? Prove that $\mathrm{MSE}(V, \theta) < \mathrm{MSE}(U, \theta)$ for all $\theta \in \Theta$.
Remark: $\mathbb{1}_{2\mathbb{N}_0}$ *is the function that maps all even numbers to 1 and all odd numbers to 0.*

Exercise 4 (**Consistency**, 4 points)**.** Consider the statistical model $\mathcal{F}_t = \{\mathrm{Unif}([0, t]) \mid t \in (0, \infty)\}$. Given an independent sample $X_1, ..., X_n$, we estimate t by

$$\hat{t}_n := \max_{i=1,\ldots,n} X_n.$$

Show that this estimator is strongly consistent.