Machine Learning (Spring 2020)	**(Due: 02.04.2020 23:55)**

Assignment 1

Instructor: Dr. Sergey Kosov *TA:* Yuhou Zhou

Course Policy: Read all the instructions below carefully before you start working on the assignment, and before you make a submission.

- The homework assignments are for practice purpose. The grade from your homework will not affect your final grade of the course.
- Please submit your answer sheet, either by a scanned copy or a typeset PDF file, to Moodle before the deadline.
- No late submission is accepted.
- You can do this assignment in groups of 2. Please submit no more than one submission per group.

Problem 1: Factorization of joint distributions (2.5+2.5=5 points)

Here we use the standard shorthand notation $P(x, y)$ for $P(X = x, Y = y)$, $P(x|y)$ for $P(X = x|Y = y)$, etc. Prove the following *factorization formula*

(a) $P(x, y, z) = P(x)P(y|x)P(z|x, y)$,

(b) $P(u|v_1, \dots, v_n) = P(u, v_1, \dots, v_n)/P(v_1, \dots, v_n)$.

Note: in statistics and machine learning in general, factorizing joint distributions into products of simpler distributions is a super common strategy.

Problem 2: Expectation, Variance, and Covariance (2.5+2.5+2.5+2.5+5+5=20 points)

(a) Use $var[f(x)] = \mathbb{E}[(f(x) - \mathbb{E}[f(x)])^2]$ show that $var[f(x)]$ satisfies $var[f(x)] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2$.

(b) Show that if two variables x and y are independent, then their covariance is zero.

(c) Suppose that the two variables x and z are statistically independent. Show that the mean and variance of their satisfies

$$\mathbb{E}[x + z] = \mathbb{E}[x] + \mathbb{E}[z]$$
$$var[x + z] = var[x] + var[z]$$

(d) Give a derivation for the formula $Cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

(e) Prove that the mean minimizes the quadratic loss, that is, for a random variable X with values in $\mathbb{R}$,

$$\mathbb{E}[X] = argmin_{x\in\mathbb{R}}(\mathbb{E}[(x - X)^2])$$

Note: *argmin* refers to the inputs, or arguments, at which the function outputs are as small as possible. e.g. $argmin_x(2x^2 + 4x - 8) = -1$

(f) Show that for two random variables X, Y with values in $\mathbb{R}$, it holds that $-1 \leq Corr(X, Y) \leq 1$. (Assuming that both RVs don't have zero standard deviation and that their joint distribution is characterized by a p.d.f. $f(x, y)$). You may use the following fact (a special case of the so-called *Cauchy-Schwarz inequality*):

$$(\int_{\mathbb{R}^2} xyf(x,y)d(x,y))^2 \leq \int_{\mathbb{R}^2} x^2 f(x,y)d(x,y) \cdot \int_{\mathbb{R}^2} y^2 f(x,y)d(x,y)$$

where f is a p.d.f. on $\mathbb{R}^2$.