On the top of your answer sheet, copy the following sentence and place your signature underneath: "*The results presented here are my own work. As instructed, I am not receiving any help from anybody nor using any literature, including course notes.*"

Write clearly and justify your answers.

Incorrect statements in essay-type questions will incur negative points.

Time limit: 90 min. The exam starts at 14.00 by publishing it through course's Slack channel. It finishes at 15.40, providing extra ten minutes for compiling the submission file and submitting it through the Slack channel's private message to the instructur (bzupan).

(1) [5 points] *(essay-type question)* We have the following learning algorithms:

- **BLR**: Bayesian linear regression with built-in standardization of non-intercept input variables and $\beta_i \sim N(0, 10)$ prior on the coefficients.
- **RF**: Random forest.
- **NN**: Feed-forward neural network trained using backpropagation of squared error.

If you think some key detail about a learner is missing, you may assume which parameter was used or how it was implemented. However, the assumption must be sensible - something that could reasonably used as a default in some implementation (do not assume behavior that makes the learner clearly behave poorly or is not even feasible to implement). And, whatever you assume, applies to all the below tasks. That is, you cannot use a different implementation or parameter values for different tasks.

Each of the tasks below require you to design a data generating process. For each task you may also specify $n \geq 10$ - the number of observations used for training. Your dataset must satisfy the task requirement in expectation over all possible training sets and assuming an ideal evaluation of true risk (out-of-sample error). Slightly less formally, it must satisfy the task requirement on average over all possible training sets, where we train the model on the training set and then evaluate its out-of-sample performance using an arbitrarily large test set.

Please specify the data generating process and any dependencies between the variables in sufficient detail to support your arguments:

a. A regression data generating process with at least one independent variable where RF will outperform BLR and NN with respect to mean squared error. Or justify why this is not possible.

b. A regression data generating process with at least one independent variable where BLR will outperform RF and NN with respect to mean squared error. Or justify why this is not possible.

c. A regression data generating process with exactly one independent variable where RF does not outperform predicting with the mean (wrt mean squared error) although the optimal model (not necessarily one of the above) can outperform predicting with the mean.

(2) [5 points] *(essay-type question)* List 3 different dimensionality reduction/clustering techniques that you know. For each of them briefly summarize in which case you would prefer over the others (or why you would never prefer to use it). For each of them determine the time complexity of applying it (how it depends on the dimensionality of the data and the number of observations).

(3) [5 points] *(essay-type question)* Joe F. Random, a machine learning practitioner, is working on a forecasting problem. The goal is for each day, given the characteristics of that day, predict a binary outcome. He has data for every day for the past 5 years. He has to deliver a model that will be used in production without any possibility of modification. He can only choose between learner A or B, both of which have some hyperparameters, and he can only use them in a black-box manner (train with some hyperparameters and use the trained model for predicting, repeating this an arbitrary number of times for different hyperparameters).

Joe decided on the following training and evaluation process. He used the first 3 years for training and hand-crafted (trial-and-error) the best hyperparameters using the fourth year as a validation set. Finally, he trained A and B, each with its chosen hyperparameters, on the first four years and evaluated them on the fifth year, selecting A which gave better results. He then trained A on all five years using the same hyperparameters and delivered to production.

a. Does Joe's model evaluation process have any potential sources of bias (positive/negative?) or variance? That is, identify all sources of error when estimating true risk (how the delivered model is expected to perform in production) with the performance on the fifth year.

b. Are there any flaws in Joe's learning process (how the models were trained and hyperparameters chosen)?

c. Suggest a better learning and evaluation process that is still within the constraints of the problem.

(4) [5 points] Some machine learning-related mathematics:

a. Show that the sum of two Mercer kernels is a Mercer kernel.

b. Show the derivation of the closed-form solution for the coefficients in ridge regression.

c. Define the Shapley-value contribution of a feature for an instance and show how a feature that has no effect on the prediction will receive a 0 Shapley-value.

The End.