🎥 Linear Algebra. Project

📅 Dates		
☑ Finished	☐	
# Module		
# Time spent, hours		
⌄ Type		

Project description

The Sure Tomorrow insurance company wants to protect its clients' data. Your task is to develop a data transforming algorithm that would make it hard to recover personal information from the transformed data. This is called **data masking,** or **data obfuscation.** You are also expected to prove that the algorithm works correctly. Additionally, the data should be protected in such a way that the quality of machine learning models doesn't suffer. You don't need to pick the best model. Follow these steps to develop a new algorithm:

- construct a theoretical proof using properties of models and the given task;

- formulate an algorithm for this proof;

- check that the algorithm is working correctly when applied to real data.

We will use a simple method of data masking, based on an invertible matrix.

Project instructions

1. Download and look into the data.

2. Provide a theoretical proof based on the equation of linear regression. The features are multiplied by an invertible matrix. Show that the quality of the model is the same for both sets of parameters: the original features and the

features after multiplication. How are the weight vectors from MSE minimums for these models related?

3. State an algorithm for data transformation to solve the task. Explain why the linear regression quality won't change based on the proof above.

4. Program your algorithm using matrix operations. Make sure that the quality of linear regression from sklearn is the same before and after transformation. Use the *R2* metric.

Data description

The dataset is stored in file `/datasets/insurance_us.csv`. <u>download dataset</u>.

- **Features:** insured person's gender, age, salary, and number of family members.

- **Target:** number of insurance benefits received by the insured person over the last five years.

Project evaluation

We've put together the evaluation criteria for the project. Read this carefully before moving on to the task.

Here's what the reviewers will look at when reviewing your project:

- Have you followed all the steps of the instructions?

- Have you justified your conclusion about matrix multiplication?

- Have you described and explained the transformation algorithm?

- Have you tested your algorithm using the data?

- Have you kept to the project structure?

- Have you kept the code neat?

You have your takeaway sheets and chapter summaries, so you are ready to proceed to the project.

Good luck!