Logistic Regression **(scored tasks: 4 points)**

Tasks (Lab 4):

1. Problem of linearly separable classes.

   Dataset *earthquake.txt.* corresponds to problem of prediction of seismic shocks (volcanic eruptions and nuclear explosions) (variable **popn**) based on two variables: **body** (deep wave magnitude) and **surface** (surface wave magnitude).

   - Make scatterplot for variables **body** and **surface**. Mark classes corresponding to observations.
   - Fit logistic model without regularization, print estimated coefficients, estimated probabilities and compute log-likelihood function.
   - Fit logistic model with ℓ_2 regularization, print estimated coefficients, estimated probabilities and compute log-likelihood function.

2. Simulation example.

   - Generate data from logistic model:

   $$y_i \sim Bern(p_i),$$

   where

   $$p_i = \frac{1}{1+\exp[-(\beta_0+\beta_1 x_{i,1}+\beta_2 x_{i,2}+\ldots+\beta_5 x_{i,5})]},$$

   for $i = 1,\ldots,n$, $x_{i,1},\ldots,x_{i,5} \sim N(0,1)$, $n = 50$. Parameters: $\beta_0 = 0.5$, $\beta_1 = \ldots = \beta_5 = 1$. Fit logistic model and calculate the estimators of the coefficients $\hat{\beta} = (\hat{\beta}_1,\ldots,\hat{\beta}_5)$. Repeat the experiment $L = 100$ times and compute the MSE (mean squared error):

   $$MSE = E(||\hat{\beta}-\beta||^2),$$

   where $||\cdot||$ is Euclidean norm, $\beta = (\beta_1,\ldots,\beta_5)$ is vector of true parameters.

   - Repeat the experiment for $n = 50, 60, 70, 80, 90, 100, 200, 300\ldots, 1000$ and make a plot showing how MSE depends on n.
   - Using the same datasets, train the model based only on 3 variables: $x_{i,1}, x_{i,2}, x_{i,3}$ and draw the analogous curve showing how MSE for $\beta = (\beta_1,\beta_2,\beta_3)$ depends on n.