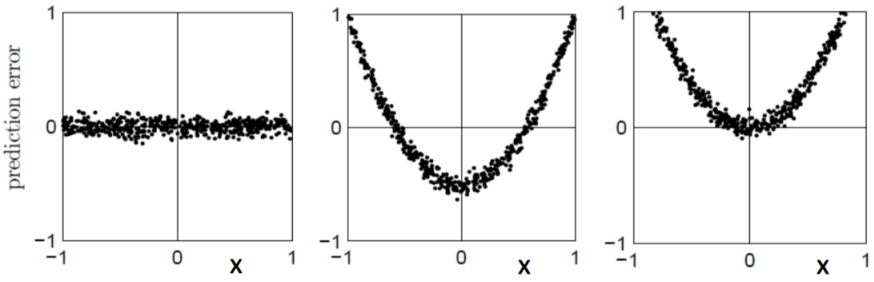
Theoretical task 3.

Recommendations: all solutions should be short, mathematically strict (unless qualitative explanation is needed), precise with respect to the stated question and clearly written. Solutions may be submitted in any readable format, including images.
Submission link: here

1. Consider linear regression task in one-dimensional space. $(x_i, y_i)_{i=1}^N$ is a training dataset and for object $i : x_i \in [-1, 1]$ is a feature, y_i is an answer we want to predict, $y_i^* = kx_i + b$ is our prediction. At the picture below you can see three different plots of the prediction error $(y - y^*)$ against x. Which of these plots cannot be obtained if least squares method is used to train a regression model?



2. Consider linear regression task in one-dimensional space $y = kx+b$ and two datasets: $(x_1^1, y_1^1), ..., (x_n^1, y_1^1)$ and $(x_1^2, y_1^2), ..., (x_m^2, y_m^2)$. Assume that the least squares method is used to train a regression models in this task.

   It turns out, that if we train a regression model on the first dataset we obtain a coefficient $k_1 > 0$. Similarly, if we train a regression model on the second dataset we obtain a coefficient $k_2 > 0$. Is it true that if we train the regression model on both datasets together then the obtained coefficient k will also be positive?

   Additionally answer that question if we know that $\sum_{i=1}^n x_i^1 = 0$ and $\sum_{i=1}^m x_i^2 = 0$.

3. Consider the dataset with the following features: height, age(from 10 to 45), sex(=1 for females, =0 for males). You would like to build a regression model to estimate person's height based on other features. Write down the regression model that would consider all the following facts:

   - on average the height of males and females is different,
   - at the age of 25 human rate of growth dramatically drops down, but still assumed to be linear.

   Explain your answer.