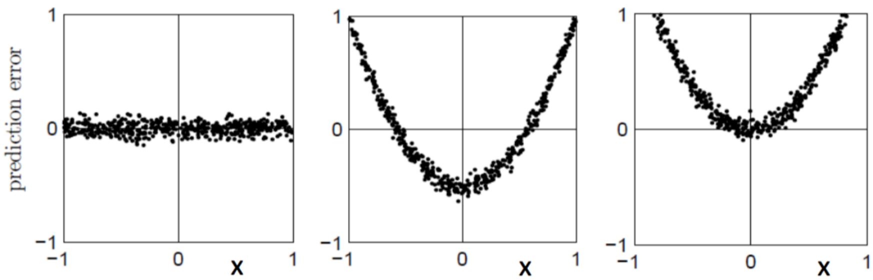
# Theoretical task 3.

*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*

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. Ridge regression is linear regression with $l_2$ regularization. Denote $X : n \times m$ as a matrix of feature descriptions of objects, $y : n$ – vector of target variable, $\beta$ – vector of weights, $\lambda$ – hyperparameter of regularization. Show, that expression of optimal $\beta$ is

$$\beta = (X^\top X + \lambda I)^{-1} X^\top y,$$

where $I$ - identity matrix of size $m \times m$

You can assume that

$$\nabla_\beta L = -2X^\top y + 2X^\top X\beta$$

and

$$\frac{dx^\top Ax}{dx} = (A + A^T)x,\ if\ A = A^T \to \frac{dx^\top Ax}{dx} = 2Ax$$

3. For linear regression task with one predictor $\hat{y} = \beta_0 + \beta_1 x$ show, that

$$\beta_1 = \frac{\sum\limits_{i=1}^n x_i y_i - \frac{\sum\limits_{i=1}^n y_i \sum\limits_{i=1}^n x_i}{n}}{\sum\limits_{i=1}^n x_i^2 - \frac{\sum\limits_{i=1}^n x_i \sum\limits_{i=1}^n x_i}{n}} = \frac{\sum\limits_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum\limits_{i=1}^n (x_i - \bar{x})^2}$$

where $\bar{x} = \frac{1}{n}\sum\limits_{i=1}^n x_i$

4. To fit linear regression we solve the following optimization task:

$$\|X\beta - y\|^2 \to \min_\beta .$$

The analytical solution of this task can be found with the following expression

$$\beta^* = (X^T X)^{-1} X^T y$$

Proof that expression using only geometric considerations. To do this, use the fact that $X\beta$ defines a linear space spanned by column vectors of $X$ (proof that $X\beta^*$ is closest vector to $y$ in linear space $X\beta$).

Hint: the shortest distance between a point and a plane is the length of a perpendicular dropped from a point to a plane.

5. 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 of 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.