theory4

by Kondrashov Anton

February 9, 2017

1 Task

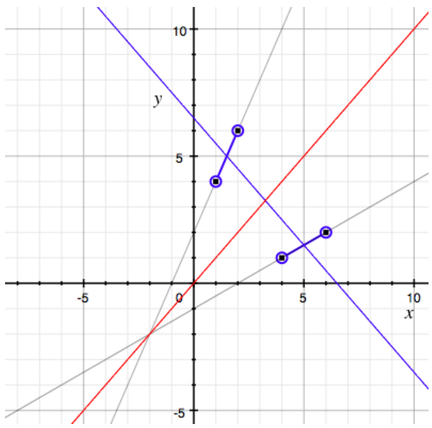
The third plot couldn't be obtained because at least one point should be under the $0x$ axis. That may be concluded from the behaviour of OLS method.
Suppose that 3 plot was built by OLS method $\Rightarrow$ All points are on the same side relative to prediction function line or on it $\Rightarrow$

1. If the line doesn't have points on it then it wasn't built using OLS method because the line that has points on it would give lesser sum of squared errors.

2. If a point lies on line then we can move the line in direction of other points in sush way that it lies between the point that lied on it before and closest point. If new sum of errors of this points is $y_1^2 + y_2^2$ while previous was $0 + (y_1 + y_2)^2$ which is more. That means that the given line wasn't built by OLS method.

2 Task

1. No it's not necessarily positive.
   Counterexample:
   First dataset $= \{(1, 4), (2, 6)\}$
   Second dataset $= \{(4, 1), (6, 2)\}$
   Then:



Points from the same dataset are connected with bold blue line. Gray lines are the ones of functions that are obtained on each separate dataset.

If we assume that the coefficient k of common prediction function is positive then optimal plot of it is red. But if we don't assume it then we see that the blue plot has the least sum of square errors on all 4 points.

2. The answer is yes it will be positive. No good idea of proof.

3 Task

$$x_0 = 1$$

$$x_1 = x_{sex}$$

$$x_3 = \begin{cases} x_{age} & \text{if } x > 25 \\ 0 & \text{if } x \leq 25 \end{cases}$$

$$x_4 = \begin{cases} 0 & \text{if } x < 25 \\ x_{age} & \text{if } x \geq 25 \end{cases}$$

$$\hat{y} = \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

I proposed x_0 to have free coefficient in function. That may be useful in general case. For x_{sex} feature I couldn't come up with better function than the one that equals the values of x_{sex} feature. Any modifications seemed senseless. I separated x_{age} feature in two features x_3 and x_4 each of which will have separate coefficient to make their influence on prediction function unequal and at the same time save linearity.