

Upper Bound Regression

Problem. Suppose we have a set of points $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ and we wish to find a line of the form

$$f(x) = \beta x + \alpha$$

such that $f(x_k) \geq y_k$ for all k and we minimize the total squared distance from each point. In other words, we want to find a line to bound above all points as tightly as possible.

This is actually useful in certain cases, and hopefully it can be used in some inferential statistics, but we shall first derive a way to calculate it. We will see that, compared to its normal linear regression equivalent, the upper bound linear regression is computed in a completely different way. Our first observation will be the following.

Observation. Suppose we have an upper bounding line $f(x)$. Then $f(x)$ must pass through at least two of $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$.