

# SML

## Practice

Given a regression train dataset  $D = \{(x, y)\} = \{(0, 1), (2, 1), (1, -1), (-1, -1), (3, -1)\}$ , find the gradient boosting tree with Huber loss. You can choose to split two different points, say at 0.5 and 2.5. Use the weak learner as decision stump. Find the prediction of sample 2.25.

- Compute  $F(x)$  as the average of all true labels. Find the residue  $y_i - F(x_i)$  for all points.
- Using  $D = \{(x, y_i - F(x_i))\}$ , evaluate which of the cuts give least SSR. That will give  $h_1(x)$ . Update the tree as  $F(x) = F(x) + 0.1h_1(x)$ . Find the residue  $y_i - F(x_i)$ .
- Using  $D = \{(x, y_i - F(x_i))\}$ , evaluate which of the cuts give least SSR. That will give  $h_1(x)$ . Update the tree as  $F(x) = F(x) + 0.1h_1(x)$ . Find the residue.
- Find  $h_2(x)$ .