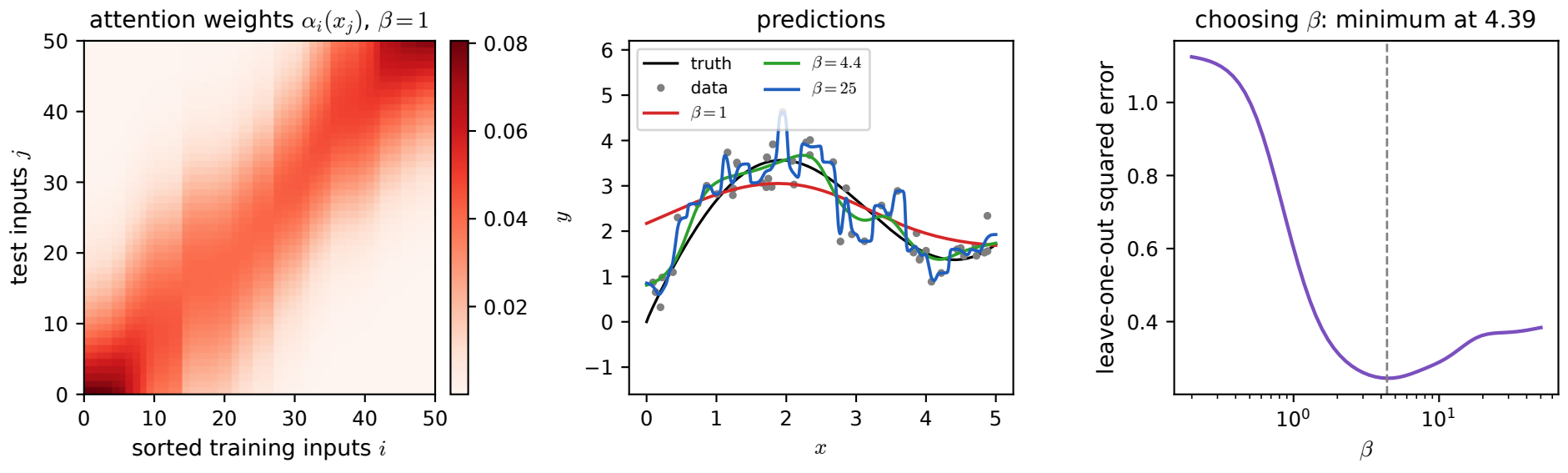

attention weights $\alpha_i(x_j)$, $\beta = 1$
test inputs j
sorted training inputs i
predictions
truth
data
$\beta = 1$
$\beta = 4.4$
$\beta = 25$
x
y
choosing β: minimum at 4.39
leave-one-out squared error
β