Homework 9

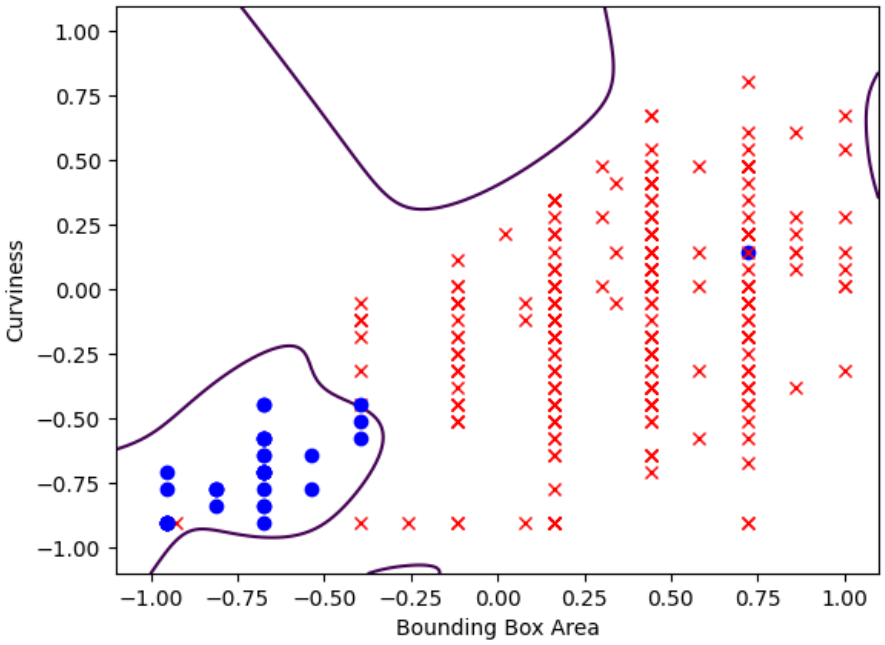
Runmin Lu

November 8, 2021

1

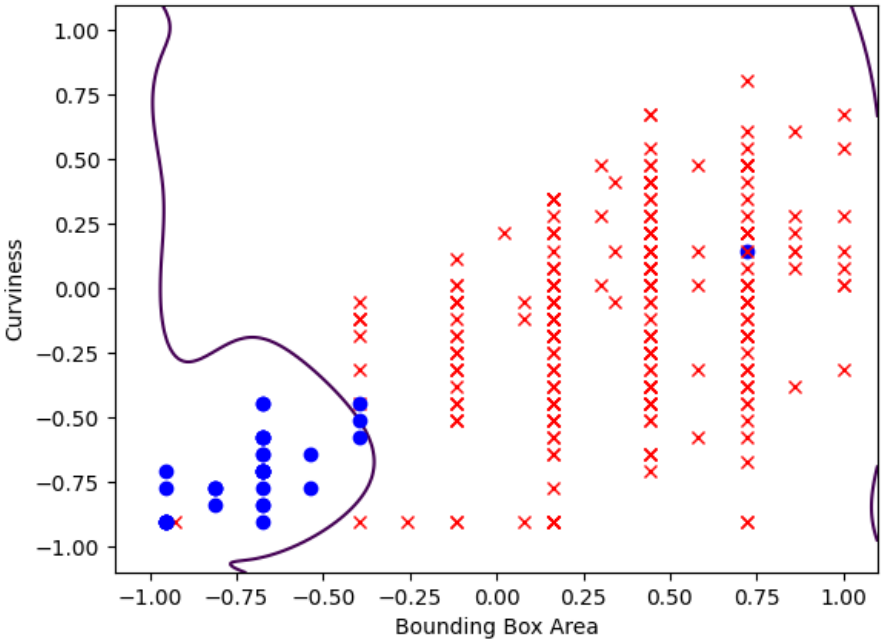
Z has dimension $N \times 45 = N \times \sum_{i=0}^{8}(i+1)$ because for each power i from 0 to 8 we have $i+1$ combinations of x and y and N is the number of data points.

2



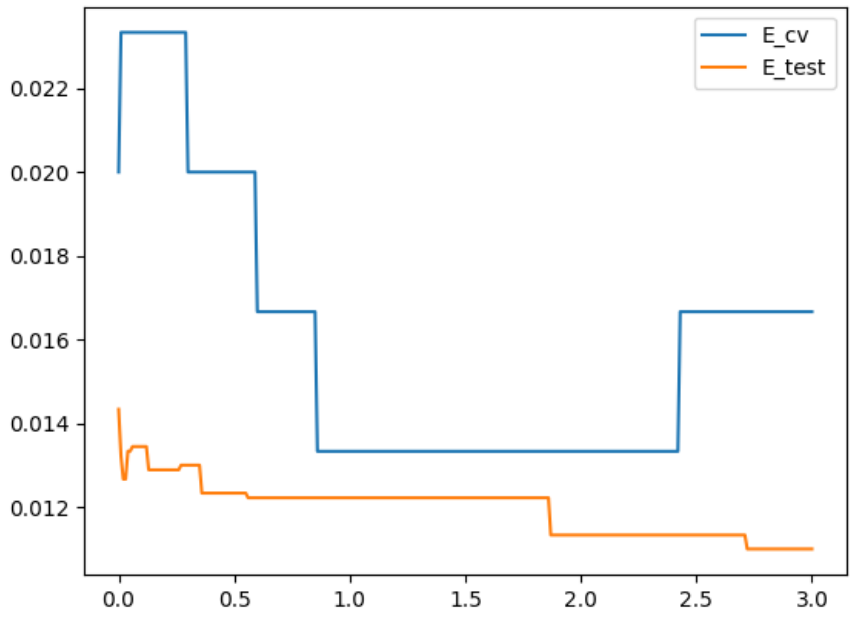
There is overfitting because the top boundary doesn't make sense. Something more curvy should not be more likely to be a 1.

3



There is still overfitting but less. The upper right boundary doesn't make sense for the same reason.

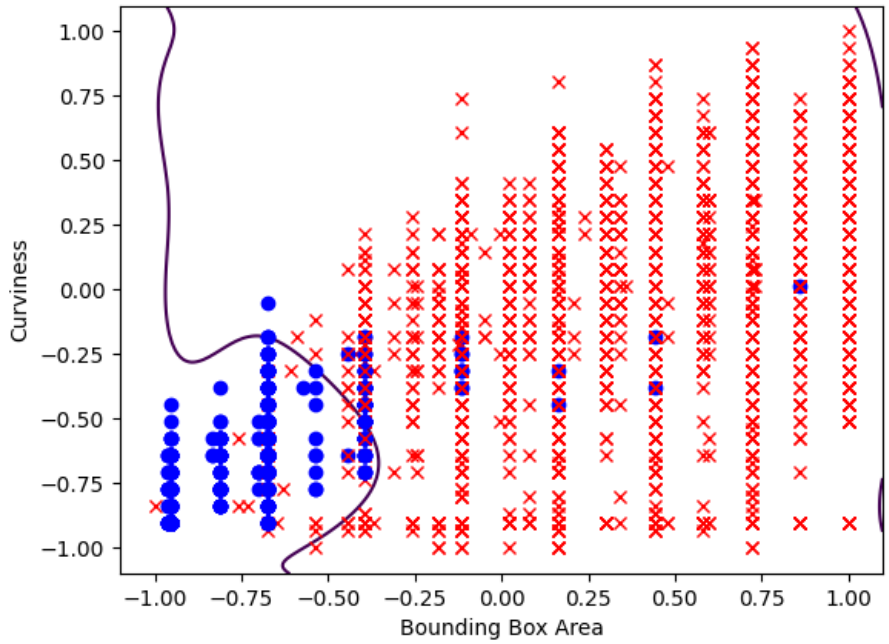
4



E_{cv} is close to E_{test}. The gap looks large because it's zoomed in and the y-axis doesn't start at 0 but if you look at the numbers, it's actually pretty small.

5

$\lambda^* = 2.42$



6

$$
\begin{aligned}
E_{\text{test}}(\mathbf{w}_{\text{reg}}(\lambda^*)) &\approx 0.0111 \\
E_{\text{out}}(\mathbf{w}_{\text{reg}}(\lambda^*)) &\leq E_{\text{test}}(\mathbf{w}_{\text{reg}}(\lambda^*)) + \sqrt{\frac{1}{2 \times 8998} \ln \frac{2}{0.01}} \\
&\approx 0.0111 + 0.0172 \\
&= 0.0283
\end{aligned}
$$

7

No because we use $E_{\text{cv}}(\lambda^*)$ to select λ^*, which is then used to select $\mathbf{w}_{\text{reg}}(\lambda^*)$ as the final hypothesis. That's where data snooping occurs. On the other hand, $E_{\text{test}}(\mathbf{w}_{\text{reg}}(\lambda^*))$ purely measures the performance of $\mathbf{w}_{\text{reg}}(\lambda^*)$ because we do not use the test set $\mathcal{D}_{\text{test}}$ in training at all.

8

Yes because there's no data snooping. $E_{\text{test}}(\mathbf{w}_{\text{reg}}(\lambda^*))$ uses $\mathcal{D}_{\text{test}}$ to evaluate the performance of $\mathbf{w}_{\text{reg}}(\lambda^*)$, which in no way affected the selection of $\mathbf{w}_{\text{reg}}(\lambda^*)$. The selection only use the 300 data points in a completely separate set $\mathcal{D}$.