Exercise 6

Machine Learning I

|  |  |
| --- | --- |
|  | 6A-1. |

First, rearrange through Bayes theorem:

Decompose:

Now we can just plug in the values from the previous exercises and slides. No need to recompute everything again:

TODO

|  |  |
| --- | --- |
|  | 6A-3. |

The noise is symmetric around . So by the same way we derived when the noise was normal, we can now say

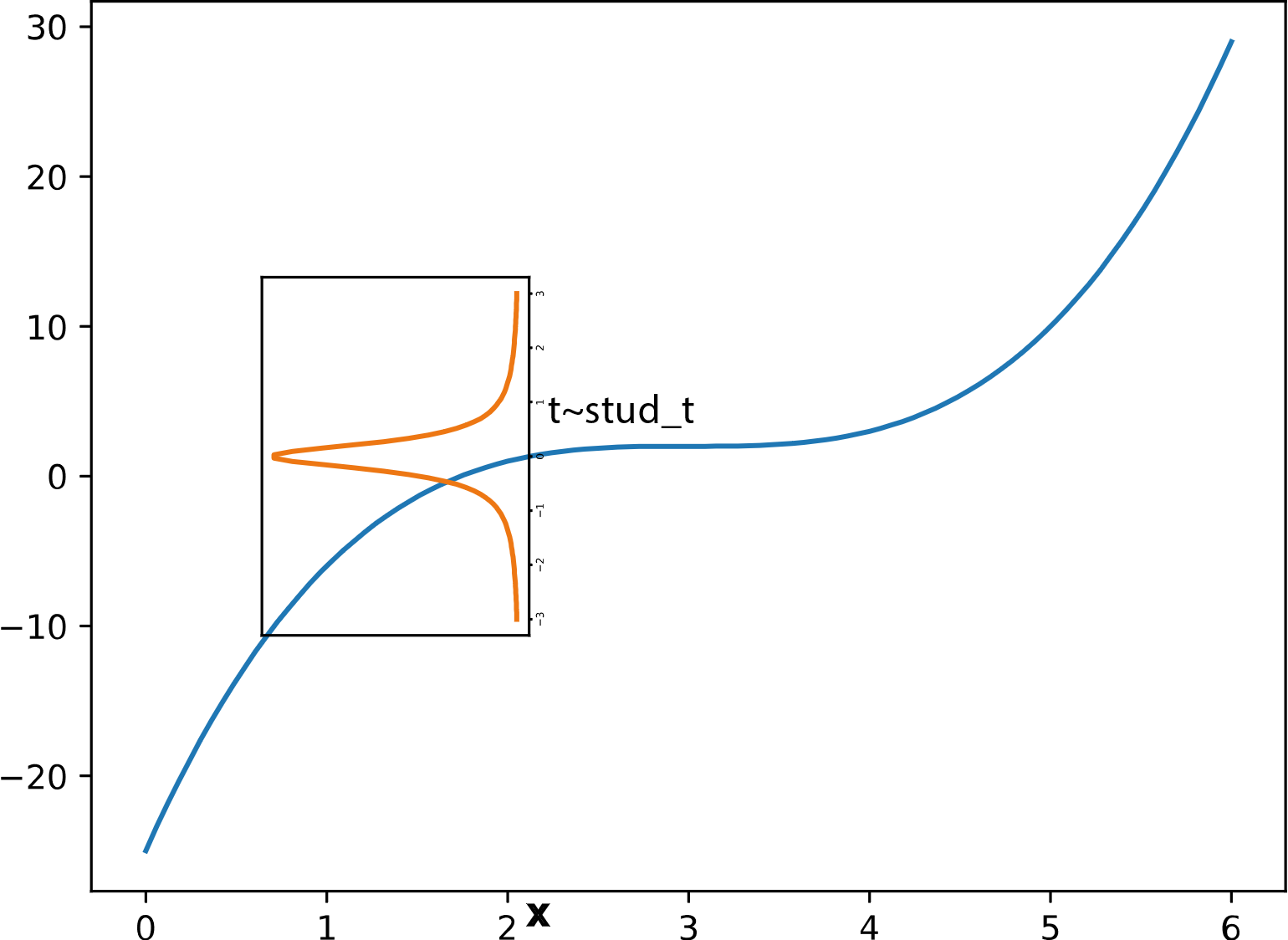


Figure 1 Because the noise is symmetrically distributed around and the   
hidden function f(x) is non-random (p(t|x)=1 for one value of t), is identically distributed  
to its noise with adjusted mean.

This leads to the log likelihood:

Let us set . We now plot the error of one datapoint in comparison to a log normal pdf. The following picture was attained with the parameters

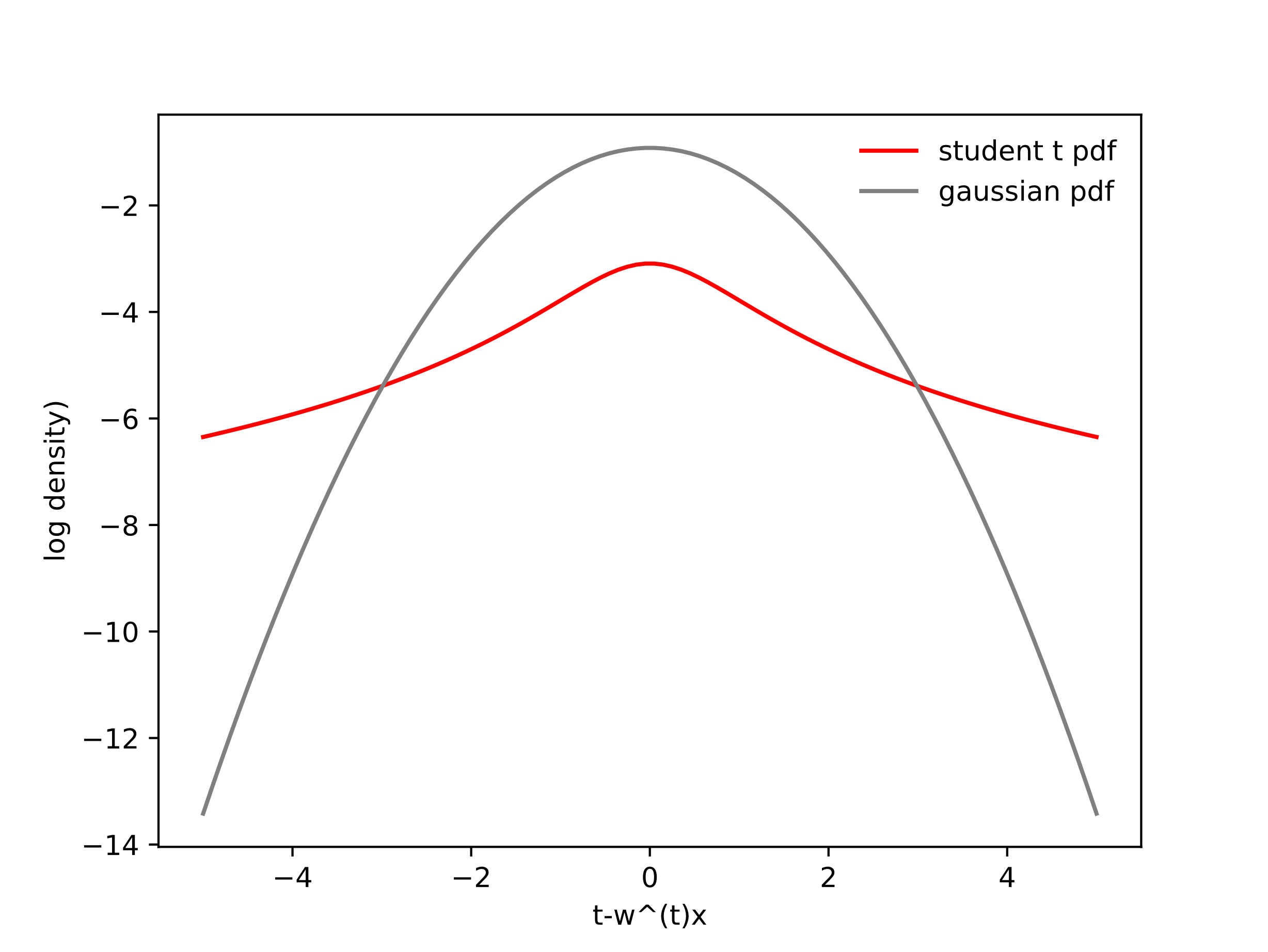


Figure 2 Comparison of the test error between a log normal pdf and  
a student t pdf. Image was attained with the parameters .

For small errors , the student t distribution appears to be less forgiving, as less density is concentrated around small errors (). Larger errors are less penalized however, as .

Generally speaking, this means if we try to maximize our weights with respect to , then our fit is trying to mitigate large errors while being mostly indifferent to smaller ones.  
If we maximize with respect to , we get a more balanced fit, we small deviations are considered.