In terms of model capacity if we only ever supply the same training dataset and we see a performance improvement by rebuilding a second or third model on the error residuals... It would seem that either the original model lacked the capacity (or complexity) to adequatly fit the data — or perhaps the model was incapable of actually representing the data set; such as when a continuous model attempts to fit a non-continuous system, such as:

discontinue

continuous.

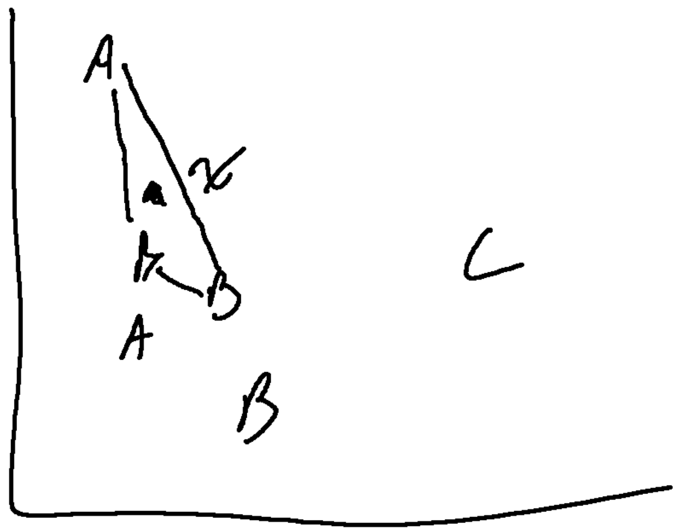
trying to get the sharpness into a non continuous region can be difficult if the model is continuous. Further confounding the issue is regularisation which may prevent a very sharp jump such as this. This is an exemplary situation where a residual prediction could iteratively add sharpness since the target is reduced in magnitude.

However the second issue becomes present in that the model is still just trying to overfit to the training data..

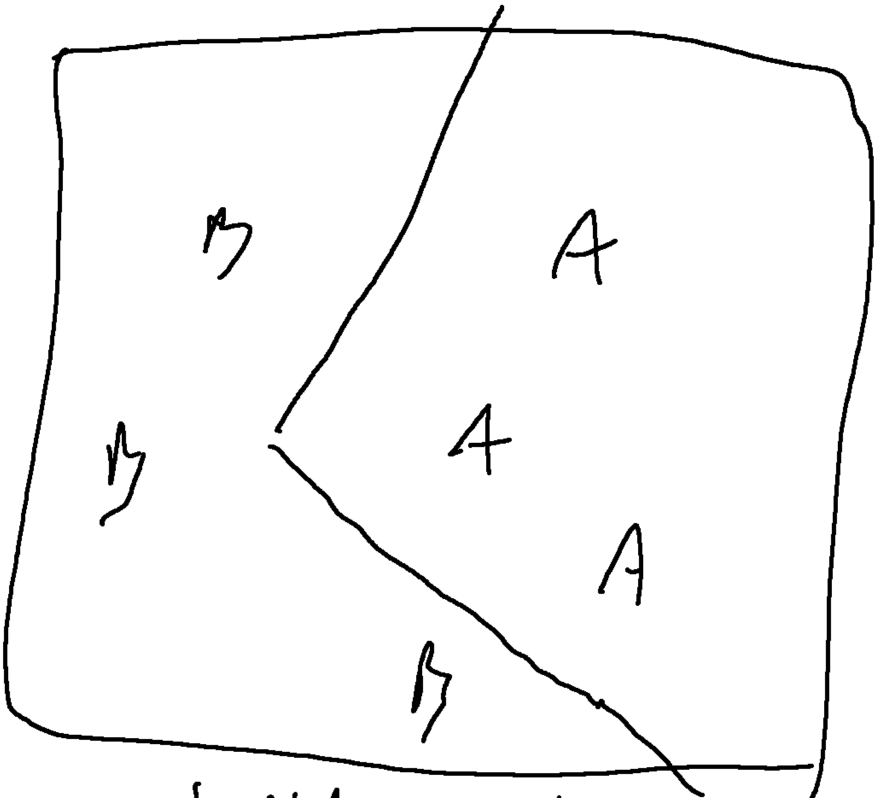
---

if we assume we have 10 classes such as in MNIST with a limited set of training data such as 1000 samples we may only have 100 of each class to learn a model from...

for a model like knn each sample stands on its own as a reference so long as we use $k=1$, however if we increase k to 2 or 3 we now run into an issue.



If we look at sample x here knowing that it is an A, with $k=3$, we don't have enough saturation to support $k=3$.
If we did something like adaboost we could give the a sample more weight.



if we build a $k=1$ voronoi this creates an optimal boundary, which is likely worth encoding... But then how do you compress this representation in a meaningful way, especially at $k=2$ or $k>1$ where the boundary is transitioning [illegible] versus flat