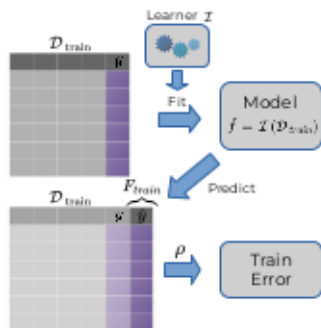


TRAINING ERROR

Simply plugin predictions for data that model has been trained on:

$$\rho(\mathbf{y}_{\text{train}}; \mathbf{F}_{\text{train}}) \text{ where } \mathbf{F}_{\text{train}} = \begin{bmatrix} \hat{f}_{\mathcal{D}_{\text{train}}}(\mathbf{x}_{\text{train}}^{(1)}) \\ \dots \\ \hat{f}_{\mathcal{D}_{\text{train}}}(\mathbf{x}_{\text{train}}^{(m)}) \end{bmatrix}$$

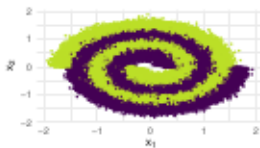


A.k.a. apparent error or resubstitution error.

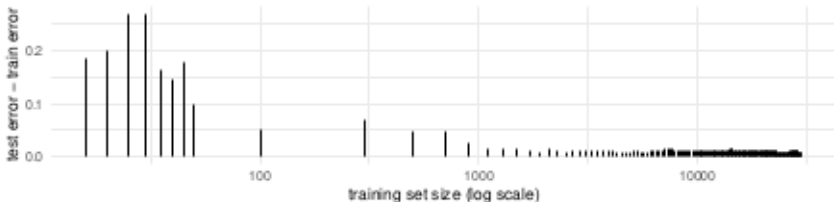
EXAMPLE 1: KNN

For large data, and some models, train error **can maybe** yield a good approximation of the GE:

- Use k -NN ($k = 15$).
- Up to 30K points from spirals to train.
- Use very large extra set for testing (to measure "true GE").

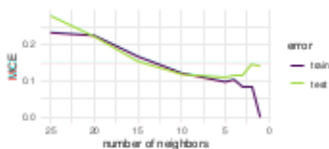


We increase train size, and see how gap between train error and GE closes.

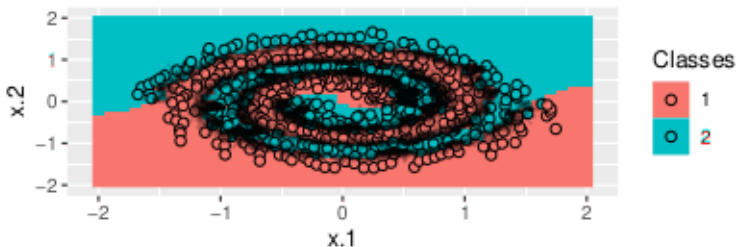


EXAMPLE 1: KNN / 2

- Fix train size to 500 and vary k .
- Low train error for small k is deceptive. Model is very local and overfits.



$k = 2$; $\text{trainerr} = 0.08$, $\text{testerr} = 0.14$

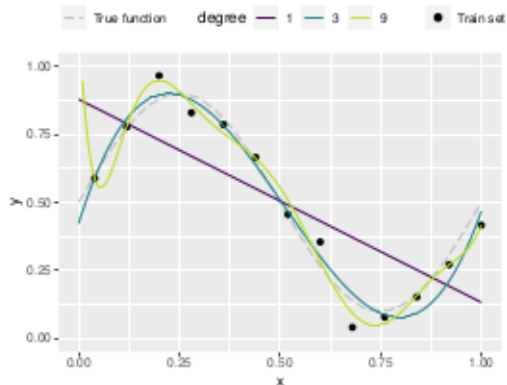


Black region are misclassifications from large test test.

EXAMPLE 2: POLYNOMIAL REGRESSION / 2

Simple model selection problem: Which d ?

Visual inspection vs quantitative MSE on training set:



- $d = 1$: MSE = 0.036:
clearly underfitting
- $d = 3$: MSE = 0.003:
pretty OK
- $d = 9$: MSE = 0.001:
clearly overfitting



Using the train error chooses overfitting model of maximal complexity.