Overfitting

Matthew Engelhard

Recall: improving one prediction may worsen another

- Suppose we predict:
  - $p_1 = .8$
  - $p_2 = .3$
  - $p_3 = .1$
- Is this a good model? Why or why not?
- Our parameters affect *all* the predictions: changing a parameter to *decrease* y_2 may also *increase* y_3

$y_1 = 1$ (dies) $y_2 = 0$ (survives) $y_3 = 0$ (survives)

p_1? x p_2? x p_3?

With a big enough model, this is no longer true.

- *perfect* predictions?

  predict $p(y_i) = y_i$

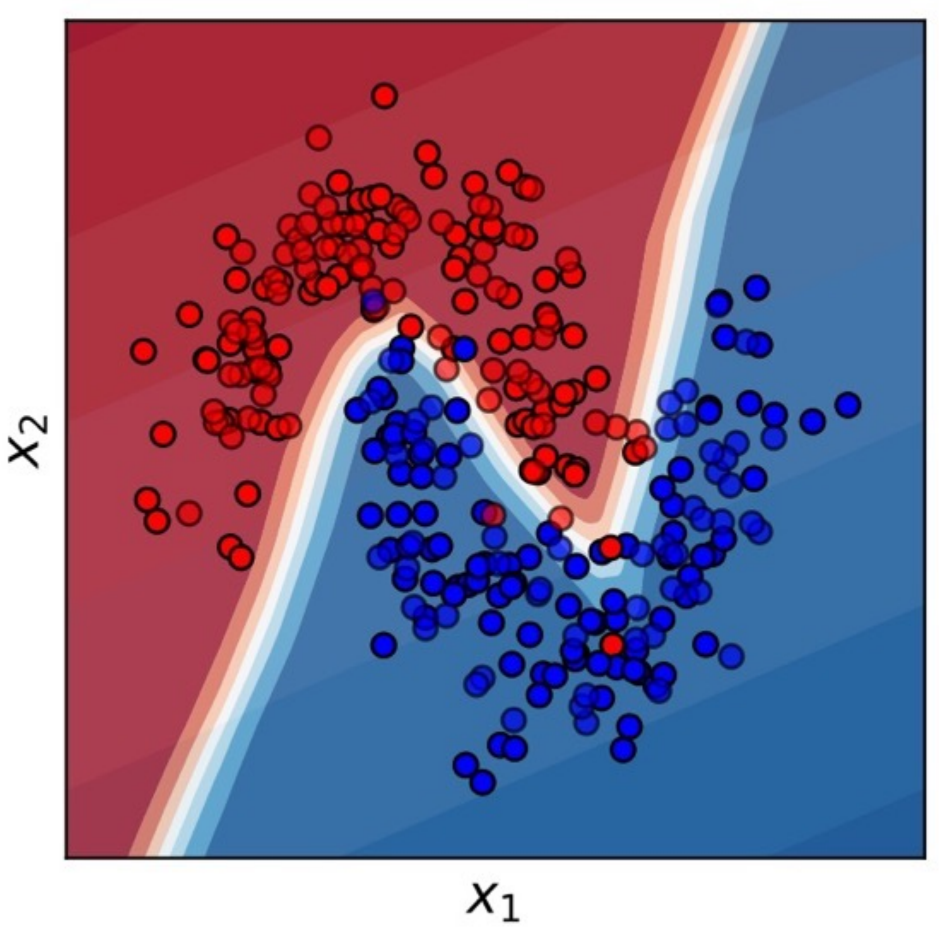
- when is this possible?

  - logistic regression where $P \gg N$
  - MLP with final hidden layer of size $M \gg N$
  - models with $\gg N$ parameters

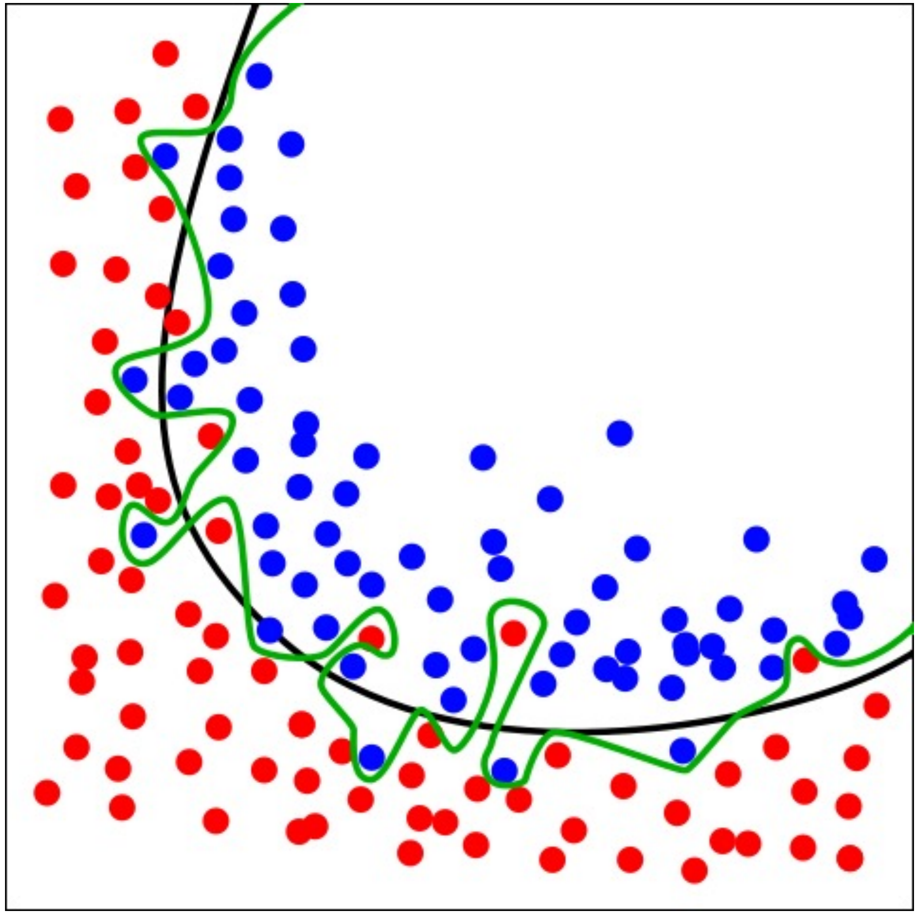
$y_1 = 1$	$y_2 = 0$	$y_3 = 0$
(dies)	(survives)	(survives)

p_1? x p_2? x p_3?

We like flexible, non-linear decision boundaries…



But when we start making our decision boundary arbitrarily complex to `fit' the training set… this is overfitting



By Chabacano - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=3610704

Green boundary:
- Correct predictions for *all* training data
- Very likely to be overfitting

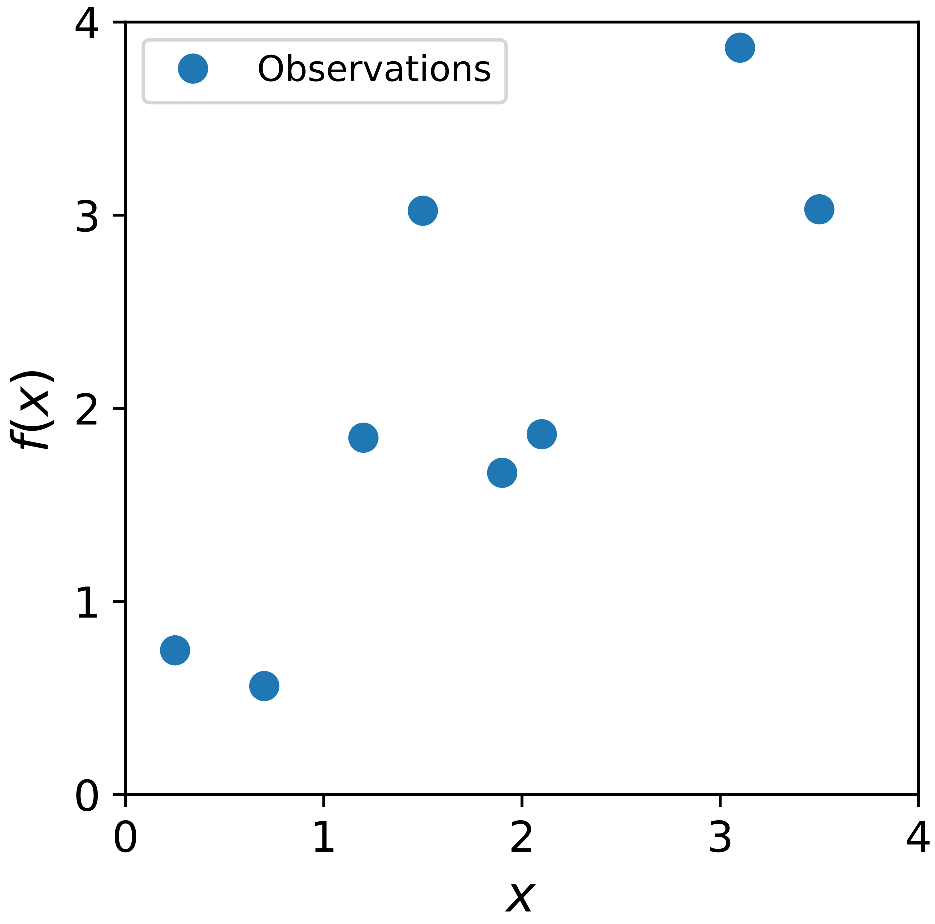
Black boundary:
- Balance between fit and model complexity

-> The black boundary is likely to perform better on new data

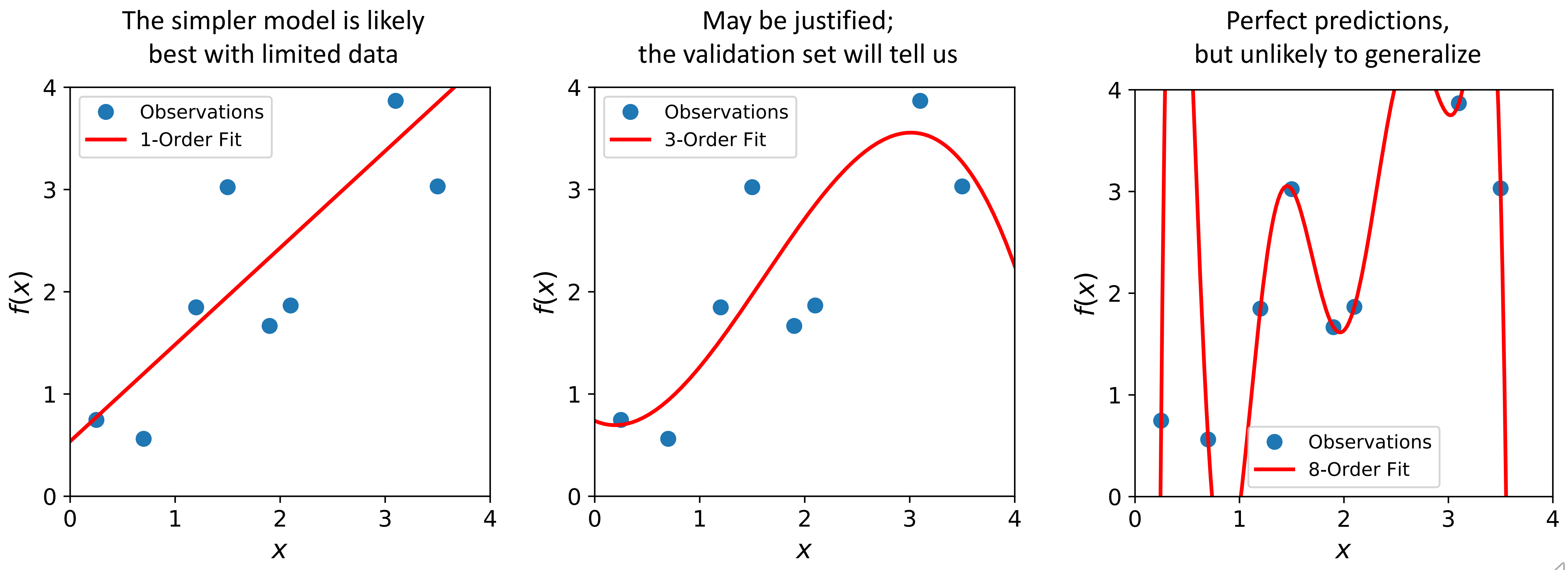
Overfitting

“Overfitting” happens when the learned model increases complexity to fit the observed training data *too well* – will not work to predict future data!

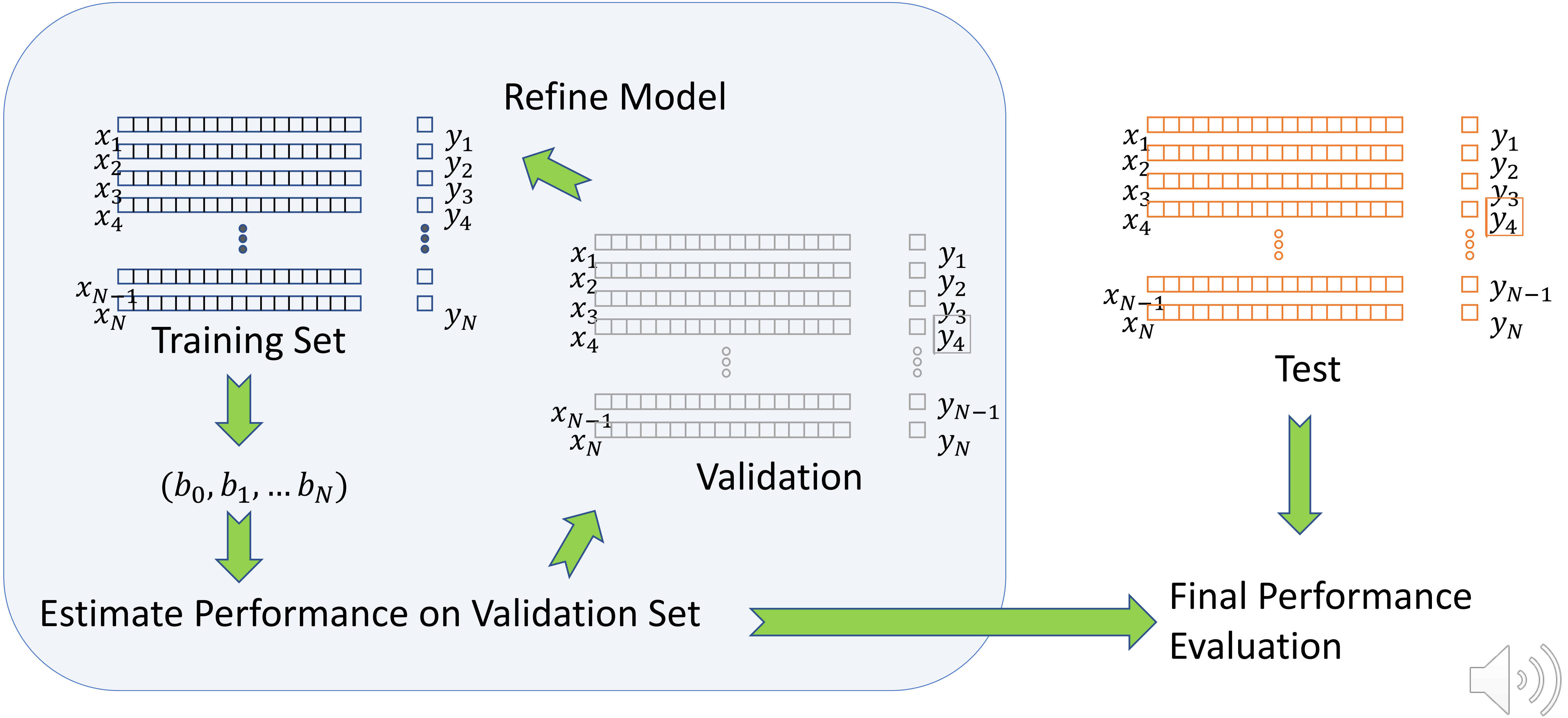
What would we want to use to fit these example data points?



With a complex enough model, we can typically predict our training labels perfectly.

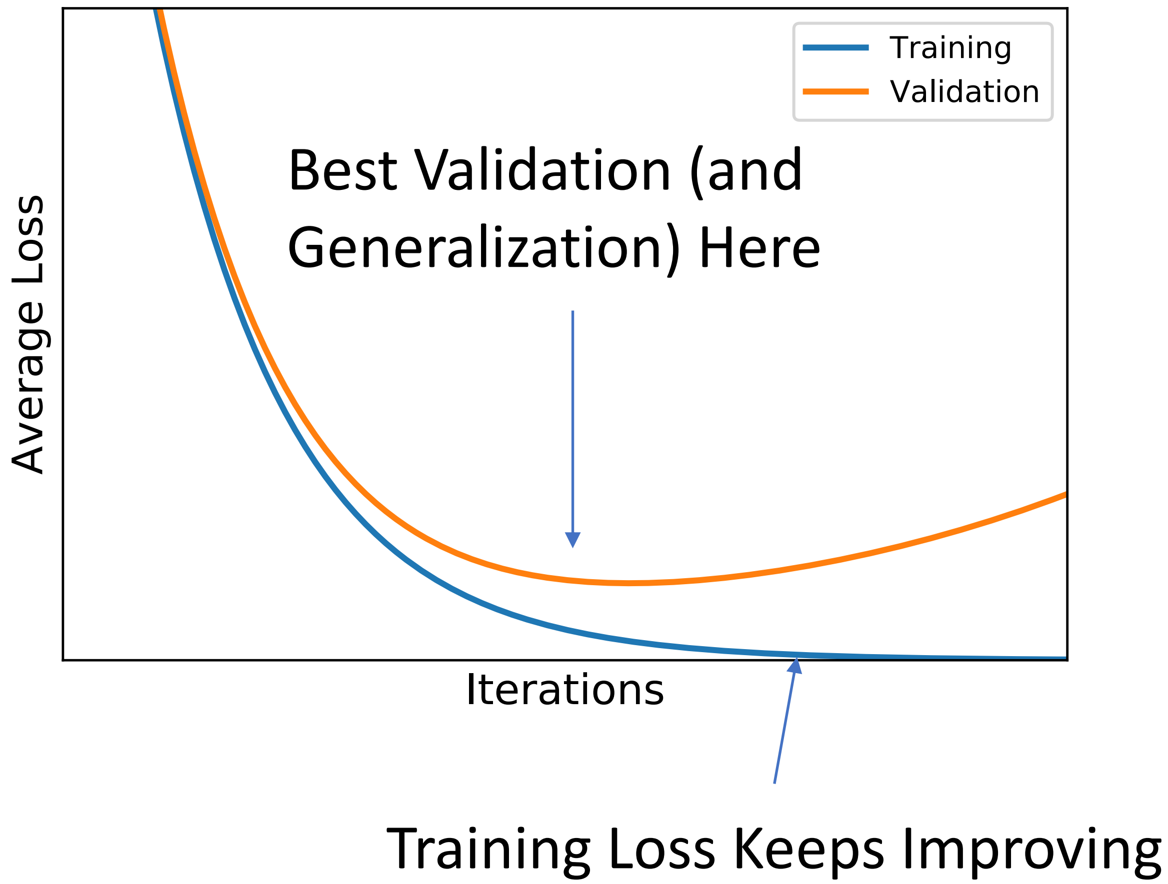


To protect against overfitting, see how well the model performs on previously unseen data.



Early Stopping

- During optimization, we can check the validation loss as we go.
- Instead of optimizing to convergence, we can optimize until the *validation* loss stops improving
  - Saves computational cost
  - Performs better on validation (and test) sets
- Widely used technique in the field



Conclusions

- Greater model complexity is often, but not always, advantageous
- Overfitting refers to learning and exploiting patterns in the training set that are not repeated in a new sample, including the validation and test sets
- Proper model validation is critical to estimate real-world performance and mitigate overfitting