

Lab 4F: Some models have curves

Directions: Record your responses to the lab questions in the spaces provided.

Making models do yoga

- Before moving on, load the movie data and split it into two sets:
- A set named `training` that includes 75% of the data.
- And a set named `testing` that includes the remaining 25%.

Problems with lines

- Train a linear model predicting `audience_rating` based on `critics_rating` for the training data. Assign this model to `movie_linear`.
- Fill in the blanks below to create a scatterplot with `audience_rating` on the y-axis and `critics_rating` on the x-axis using your testing data.
- Run the code below to add the *line of best fit* for the training data to the plot.

- Describe, in words, how the line fits the data. Are there any values for `critics_rating` that would make obviously poor predictions?
- Compute the MSE of the model for the testing data and write it down for later.

Adding flexibility

Making bend-y models

- Fill in the blanks below to train a quadratic model predicting `audience_rating` from `critics_rating`, and assign that model to `movie_quad`.
- What is the role of the number 2 in the `poly()` function?

Comparing lines and curves

- Fill in the blanks below to
- create a scatterplot with `audience_rating` on the y-axis and `critics_rating` on the x-axis using your testing data, and

- add the *line of best fit* and *best fitting quadratic curve*.
- Compare how the *line of best fit* and the *quadratic* model fit the data. Which do you think has a lower test MSE?
- Compute the MSE of the quadratic model for the test data and write it down for later.
- Use the difference in each model's test MSE to describe why one model fits better than the other.

On your own

- Create a model that predicts `audience_rating` using a cubic curve (polynomial with degree 3), and assign this model to `movie_cubic`.
- Create a scatterplot with `audience_rating` on the y-axis and `critics_rating` on the x-axis using your test data.

- Using the names of the three models you have trained, add the *line of best fit*, *best fitting quadratic curve*, and *best fitting cubic curve* for the training data to the plot.
- Based on the plot, which model do you think is the best at predicting the testing data?
- Use the difference in testing MSE to verify which model is the best at predicting the testing data.