

Name: _____

Date: _____

LAB 4E: Some Models Have Curves

Response Sheet

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

Making models do yoga

(1) Before moving on, load the movie data and write and run code splitting it into two sets:

- A set named `training` that includes 75% of the data.
- And a set named `test` that includes the remaining 25%.
- Remember to use `set.seed`.

Problems with lines

(2) Write and run code training a linear model predicting `audience_rating` based on `critics_rating` for the training data. Assign this model to `movie_linear`.

(3) 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 test data.

`xyplot(_____ ~ _____, data = _____)`

(4) Describe, in words, how the line fits the data. Are there any values for `critics_rating` that would make obviously poor predictions?

(5) Compute the MSE of the model for the test data and write it down for later.

Adding flexibility

Making bend-y models

(6) Fill in the blanks below to train a quadratic model predicting `audience_rating` from `critics_rating`, and assign that model to `movie_quad`.

`movie_quad <- lm(_____ ~ poly(_____, 2), data = training)`

(7) What is the role of the number 2 in the `poly()` function?

Name: _____

Date: _____

LAB 4E: Some Models Have Curves

Response Sheet

Comparing lines and curves

(8) Fill in the blanks to

- create a scatterplot with `audience_rating` on the y-axis and `critics_rating` on the x-axis using your test data, and
- add the *line of best fit* and *best fitting quadratic curve*.

```
xyplot(_____ ~ _____, data = _____)
```

```
add_curve(_____, col = "blue")
```

```
add_curve(_____, col = "red")
```

(9) Compare how the *line of best fit* and the *quadratic* model fit the data. Which do you think has a lower test MSE?

(10) Compute the MSE of the quadratic model for the test data and write it down for later.

(11) Use the test MSE to explain why one model fits better than the other.

On your own

(12) Write and run code creating a model that predicts `audience_rating` using a cubic curve (polynomial with degree 3), and assign this model to `movie_cubic`.

(13) Write and run code creating a scatterplot with `audience_rating` on the y-axis and `critics_rating` on the x-axis using your test data.

(14) Based on the plot, which model do you think is the best at predicting the test data?

(15) Use the test MSE to verify which model is the best at predicting the test data.