Predictive Modeling

Chapter 5: Measuring Performance in Regression Models

STA 6543

The University of Texas at San Antonio

Outline

- Quantitative measures of performance
  - The mean absolute error (MAE)
  - The mean squared error (MSE)
  - The root mean squared error (RMSE) $= \sqrt{MSE}$
  - The coefficient of determination (R^2) $\rightarrow [\text{correlation}(y, \hat{y})]^2$
- The bias-variance trade-off
- Training MSE vs. Test MSE

Quantitative measures of performance

response

- When the outcome is a *numerical* value, we often measure the performance of a predictive model in terms of
  - The mean absolute error (MAE)

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

where y_i is the outcome and $\hat{y}_i$ is its predicted value from the predictive model.

  - The mean squared error (MSE)

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$RMSE = \sqrt{MSE}$

  - The root mean squared error (RMSE)
  - The coefficient of determination (R^2)

Quantitative measures of performance

```
#install.packages('Metrics')
library(Metrics)

# Use the 'c' function to combine numbers into a vector
y <- c(0.22, 0.83, -0.12, 0.89, -0.23, -1.30, -0.15, -1.4,
0.62, 0.99, -0.18, 0.32, 0.34, -0.30, 0.04, -0.87,
0.55, -1.30, -1.15, 0.20) #observed

yhat <- c(0.24, 0.78, -0.66, 0.53, 0.70, -0.75, -0.41, -0.43,
0.49, 0.79, -1.19, 0.06, 0.75, -0.07, 0.43, -0.42,
-0.25, -0.64, -1.26, -0.07)  #predicted
```

outcome (response)

($\hat{y}$, predicted value from a predictive model)

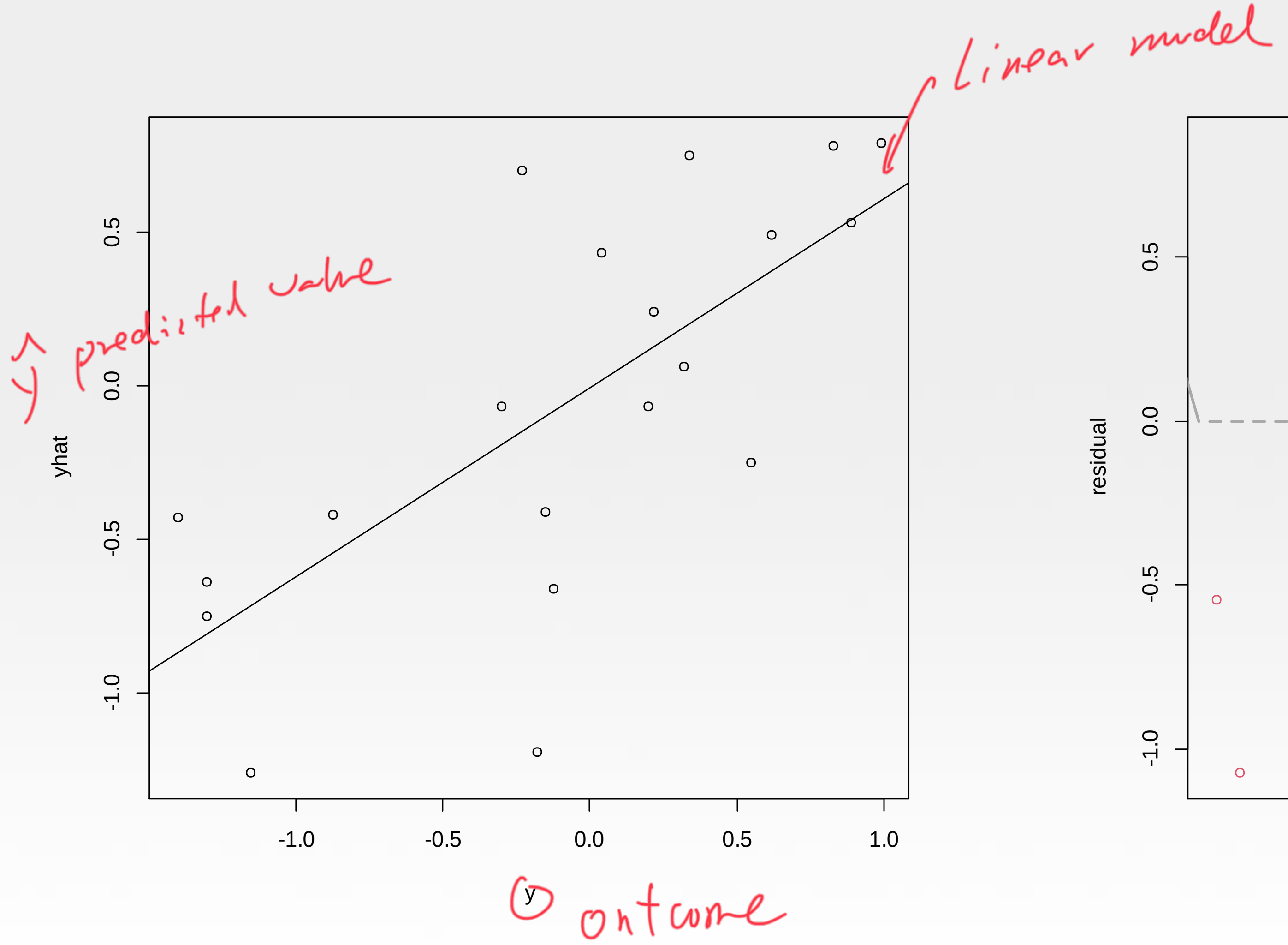
Visualize the results

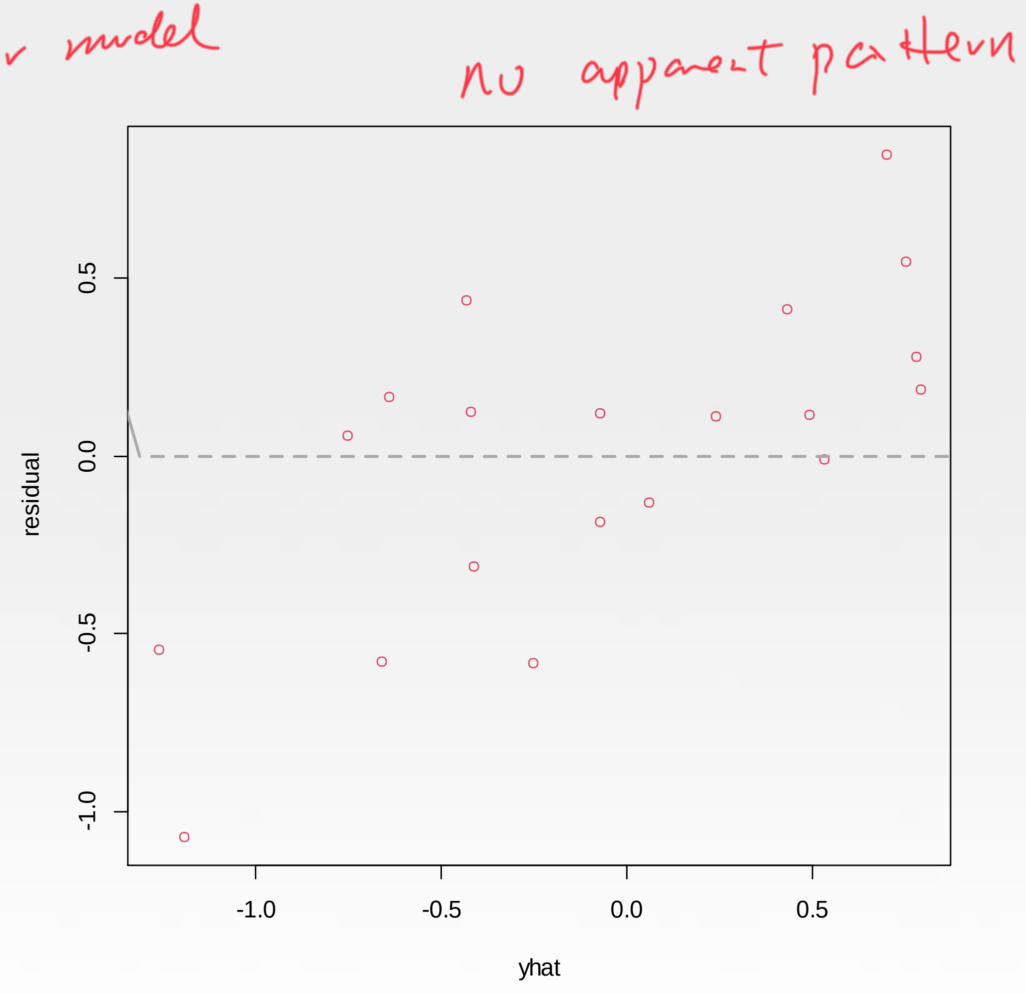
```
#Visualize the results
diff <- y - yhat

#plot of the observed and predicted values
plot(y, yhat)
fit = lm(yhat~y)
abline(fit)

#plot of the predicted values and residuals
residual = resid(fit)
plot(yhat, residual, col=2)
abline(h = 0, col = "darkgrey", lty = 2, lwd=2)
```

Visualize the results





```
#MAE
mae(y, yhat)
#or
mean(abs(y - yhat))

#MSE
(MSE = mean((y - yhat)^2))

#RMSE
sqrt(MSE)
#or
RMSE(y, yhat)

#R^2
cor(y, yhat)^2
#or
R2(y, yhat)
```

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

$$MSE = \frac{1}{n^2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\sqrt{MSE}$$

$$r : \to r^2 = R^2$$

R results

```
> #MAE
> mae(y, yhat)
[1] 0.43
> #or
> mean(abs(y - yhat))
[1] 0.43
>
> #MSE
> (MSE = mean((y - yhat)^2))
[1] 0.27404
>
> #RMSE
> sqrt(MSE)
[1] 0.5234883
> #or
> RMSE(y, yhat)
[1] 0.5234883
>
>
> #R^2
> cor(y, yhat)^2
[1] 0.5170123
> #or
> R2(y, yhat)
[1] 0.5170123
```

- MAE and RMSE — Which Metric is Better? https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-better-e60ac3bde13d

The bias-variance trade-off

E[X̄] = μ

- If we assume that the data are statistically independent and that the residuals have a theoretical mean of zero and a constant variance of σ^2, then

noise

Expected value

$$E(\text{MSE}) = \sigma^2 + (\text{Model Bias})^2 + \text{Model Variance} = \text{Constant}$$

- σ^2 is usually called "irreducible noise" and cannot be eliminated by modeling.

The bias-variance trade-off

- **Bias** refers to the error that is introduced by modeling a real-life problem (that is usually extremely complicated) by a much simpler problem.
  - For example, linear regression assumes that there is a linear relationship between *Y* and *X*. It is unlikely that, in real life, the relationship is exactly linear so some bias will be present.
  - The more flexible/complex a method is the less bias it will generally have.

KNN(1) → traing error is zero

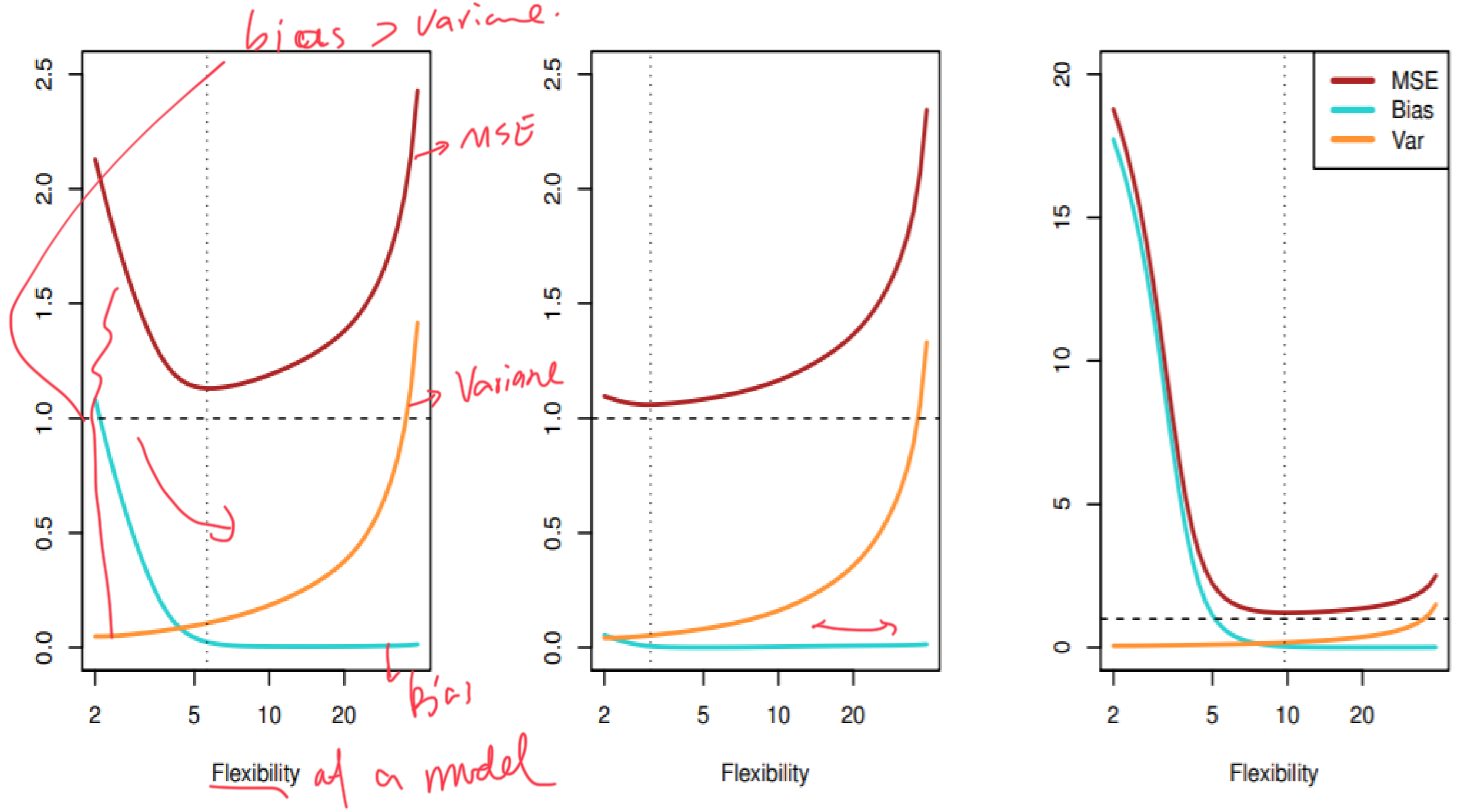
KNN(100)

The variance-bias trade-off

- **Variance** refers to how much your estimate for the predictive model would change by if you had a different training data set.
  - If method has high variance, then small changes in the training data can result in large changes in the estimated model.
  - Generally, the more flexible a method is, the more variance it has

more flexible of a method
bias ↓
variance ↑

Three examples for the bias-variance trade-off



As the model becomes more flexible, bias usually decreases, Variance usually goes up!

MSE will be U-shaped.

The bias-variance trade-off

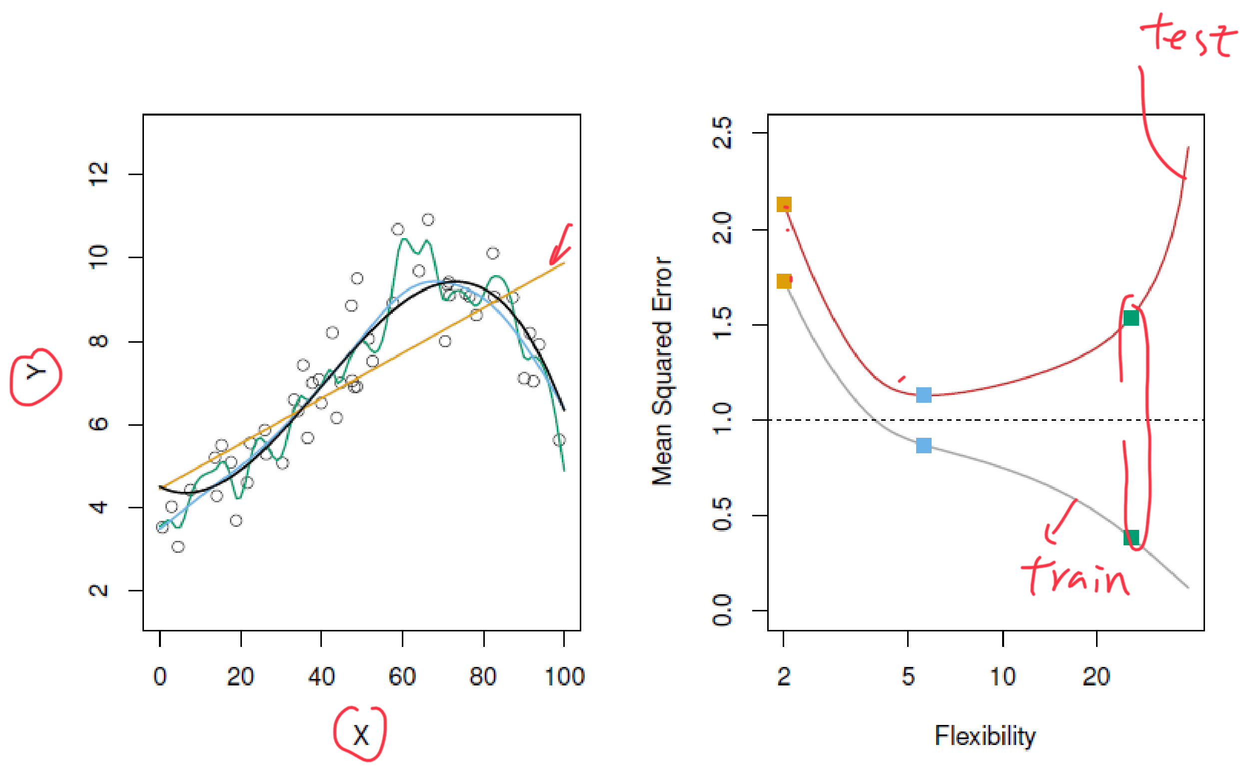
- The trade-off means that as a predictive model gets more flexible, the bias will decrease, and the variance will increase but expected test MSE may go up or down. depending on whether bias or variance dominates

Training MSE vs. Test MSE

- In general, the more flexible (i.e., more model parameters) a method is, the lower its training MSE (apparent error)will be, that is, it will fit or explain the training data very well.
  (KNN 1)
- However, the test MSE may in fact be higher for a more flexible method than for a simple approach like linear regression.
  - Recall: More flexible methods can generate a wider range of possible models as compared to less flexible and more restrictive methods (such as linear regression). The less flexible the method, the easier to interpret the model. Thus, there is a trade-off between flexibility and model interpretability.

An illustrative example

- LEFT: Black—truth. Orange—linear fit. Blue—smoothing spline. Green—smoothing spline (higher level flexibility).
- RIGHT: Red—test MSE. Grey—training MSE. Dashed—minimum possible test MSE (irreducible error).



As model becomes more flexible, traing MSE ↓, test ↓ ↑

A fundamental picture

- Training errors will always decline in general.
- However, test errors will decline at first (as reductions in bias dominate) but will then start to increase again (as increases in variance dominate).
- Keep this picture in mind when choosing a learning method. More flexible/complicated is not always better.