

Chapter 5

Chapter 5. Measuring performance in regression models

5.1 Quantitative measures of performance

RMSE - Root Mean Squared Error: How far on average the residuals are from zero or as the average distance between the observed values and the model predictions

R²: Proportion of the information in the data that is explained by the model (R² is a measure of correlation, not accuracy) (R² is dependent on the variation in the outcome)

5.2 The variance-bias trade-off

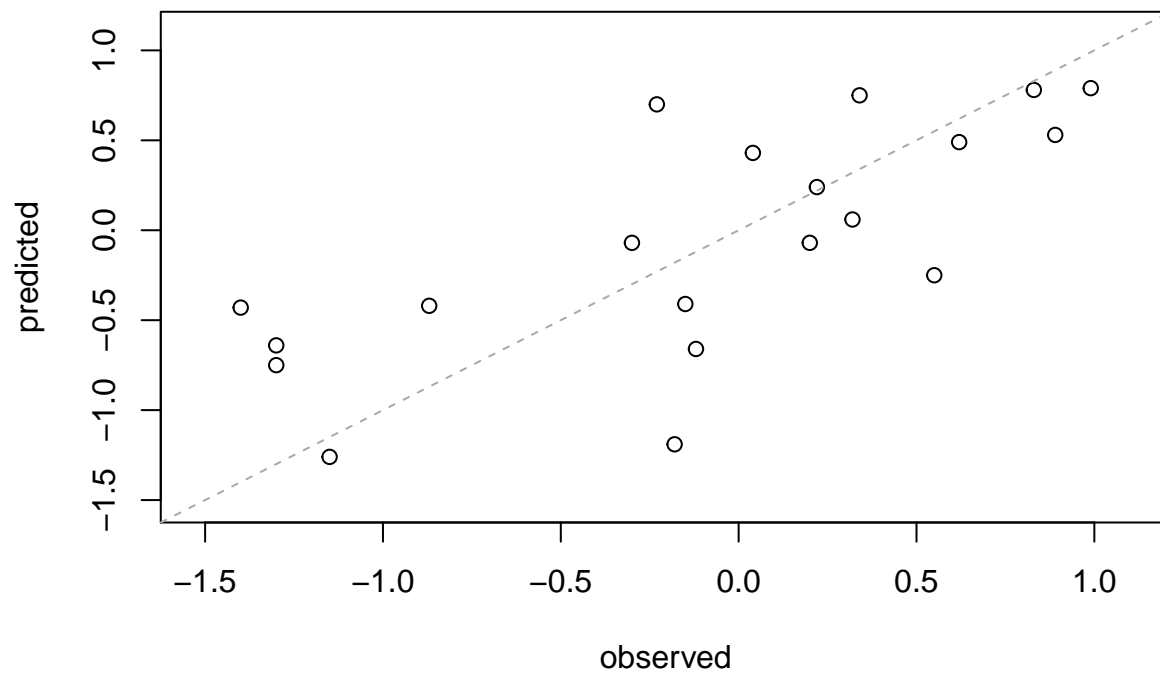
$E[\text{MSE}] = \text{irreducible noise} + \text{squared bias} + \text{variance}$

5.3 Computing

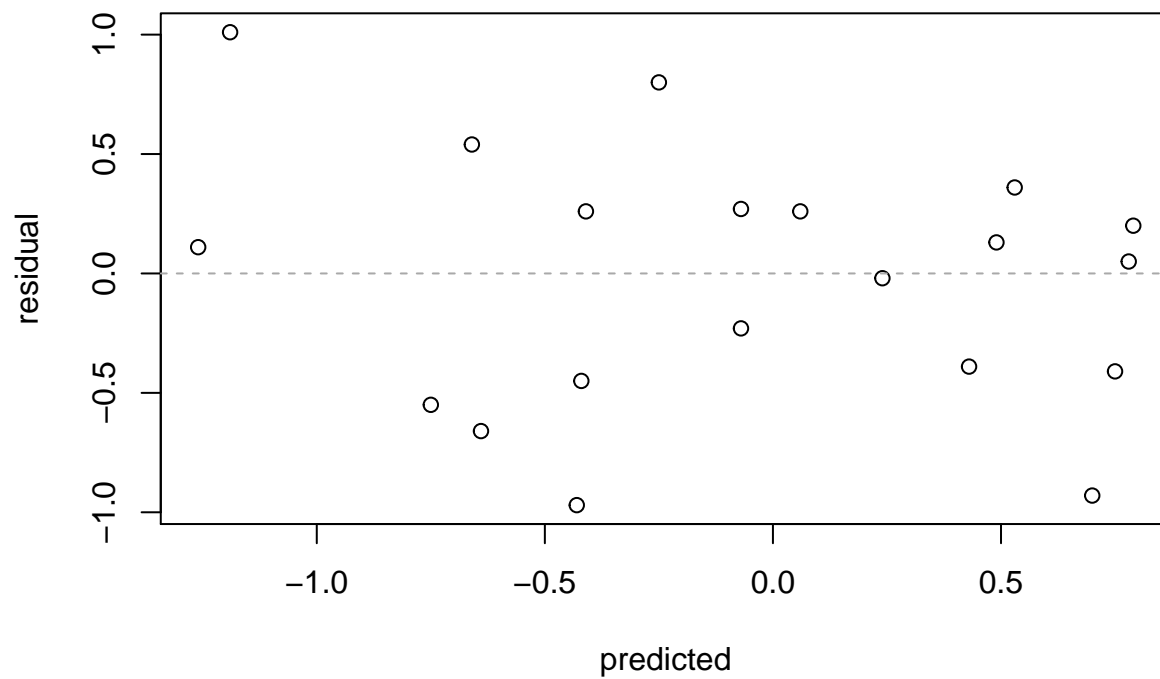
```
observed <- c(0.22, 0.83, -0.12, 0.89, -0.23, -1.30, -0.15, -1.4, 0.62, 0.99, -0.18, 0.32, 0.34, -0.30,
predicted <- c(0.24, 0.78, -0.66, 0.53, 0.70, -0.75, -0.41, -0.43, 0.49, 0.79, -1.19, 0.06, 0.75, -0.07
residualValues <- observed - predicted
summary(residualValues)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -0.9700 -0.4200  0.0800 -0.0310  0.2625   1.0100
```

```
axisRange <- extendrange(c(observed, predicted))
plot(observed, predicted, ylim = axisRange, xlim = axisRange)
abline(0, 1, col = "darkgrey", lty = 2)
```



```
plot(predicted, residualValues, ylab = "residual")
abline(h = 0, col = "darkgrey", lty = 2)
```



```
caret::R2(predicted, observed)
```

```
## [1] 0.5170123
```

```
var(predicted, observed)
```

```
## [1] 0.3525263
```

```
caret::RMSE(predicted, observed)
```

```
## [1] 0.5234883
```

```
cor(predicted, observed)
```

```
## [1] 0.7190357
```

```
cor(predicted, observed, method = "spearman")
```

```
## [1] 0.7554552
```