Maximum likelihood estimation and resampling techniques

Maximum likelihood estimation in linear regression

1. Write up the simple linear regression model with one explanatory variable.
2. Write up an expression for the log-likelihood for `n` independent observations from the simple linear regression model.
3. Show that the maximum likelihood estimators (MLE) for the intercept α and slope β are indeed the least squares estimators.

Consider a simple linear regression model for the `trees` dataset with `Volume` as response variable and `Girth` as explanatory variable.

4. Find the MLE estimates for α, β and σ numerically by maximizing the log-likelihood function.
5. Find the least squares estimates of α and β numerically. That is, define a function that computes the least squares as a function of α and β and use `optim()` to minimize the function. Compare with 4.
6. Compare 4. and 5. to the output of `summary(lm(...))`.

Overfitting and cross validation

We consider the `Credit` dataset from the `ISLR` package which contains data about credit card holders.

```
library(ISLR) # Remember that this package must be installed
head(Credit)
```

```
##   ID  Income Limit Rating Cards Age Education Gender Student Married Ethnicity
## 1  1  14.891  3606    283     2  34        11   Male      No     Yes Caucasian
## 2  2 106.025  6645    483     3  82        15 Female     Yes     Yes     Asian
## 3  3 104.593  7075    514     4  71        11   Male      No      No     Asian
## 4  4 148.924  9504    681     3  36        11 Female      No      No     Asian
## 5  5  55.882  4897    357     2  68        16   Male      No     Yes Caucasian
## 6  6  80.180  8047    569     4  77        10   Male      No      No Caucasian
##   Balance
## 1     333
## 2     903
## 3     580
## 4     964
## 5     331
## 6    1151
```

Our response variable will be `Balance` which is the constumers credit card debt. As predictor we use the variable `Rating` which is the costumer's credit rating.

1. Fit a linear regression model for the relationship between `balance` and `Rating`.
2. Use bootstrap to estimate the standard errors of the parameter estimates in the simple linear regression model. Compare to those obtained from `summary()`.

3. Use resampling of residuals to estimate the standard errors of the parameter estimates in the simple linear regression model. Compare to those obtained from `summary()`.
4. Use cross validation to decide between linear and polynomial regression.