Model evaluation

Dominik Klepl

12/5/2019

According to the model selection the best model structure is: y ~ x1 + x2 + x4 + error. Now we're going to test the model further.

Fit the model

```
data = read.csv("data/x_y.csv", header = F)
colnames(data) = c("x", "y")
```

Construct X matrix

```
X = cbind(data$x,
          data$x^2,
          data$x^4
          )

colnames(X) = c("x1", "x2", "x4")
```

Fit model, generate predictions, compute error (SSE)

```
theta = solve(crossprod(X), crossprod(X, data$y))

y_pred = X %*% theta

residuals = (data$y - y_pred)
SSE = sum(residuals^2)

#calculate prediction CI
sigma_sq = SSE/(nrow(X) - 1)
```

Parameter covariance matrix

```
cov = sigma_sq * (solve(t(X) %*% X))
print(cov)
```

```
##                x1            x2            x4
## x1  3.841759e-05  4.010982e-06 -1.065549e-06
## x2  4.010982e-06  3.450809e-05 -4.194641e-06
## x4 -1.065549e-06 -4.194641e-06  7.424601e-07
```

Plot parameter uncertainty pdf

Because we have 3 parameters, we'll need to plot all their combinations resulting in 3 plots.

First we create grid of possible parameter values for which we estimate the uncertainty. Because the estimates should come from normal distribution we'll create grids in range estimate +- 2SD.