|  |  |  |  |
| --- | --- | --- | --- |
|  | Model | Parameters | MSE |
| Regression | Linear | Best subset (nb of predictors: 6) | 0.47 |
| Cross validation + Ridge (Lambda = 0.15) | 0.42 |
| Cross validation + Lasso (lambda = 0.02) | 0.43 |
| Polynomial | Degree 3 | 0.39 |
| Degree 7 | 0.35 |
| Degree 9 | 0.321 |
| Splines | Splines (poly 9, 14 degrees of freedom) | 0.319 |
| Natural splines (poly 3, 15 degrees of freedom) | 0.37 |
| GAM (poly 3, 12 degrees of freedom) | 0.35 |
| Class. | Logistic regression | Cross validation + Best subset + bootstrap | 80.5%  accuracy |

The best regression model to predict the weight of a baby is the splines with 9 degrees of polynomial and 14 degrees of freedom. With this model, we obtain a Mean squared error 0.321 on the test set.

The best classification model to predict if a baby’s weight will be below the threshold of 2.5 kilograms is a logistic regression using cross validation, best subset and boostrap. With this model, we obtain an accuracy 80.5% on the test set.