**Section 6.5: Linear Model Selection and Regularisation in Python**

**Duration:** 3 hours

**Concepts:**

* Best subset selection
* Stepwise selection
* Ridge regression
* The lasso

**Textbook section:** An Introduction to Statistical Learning in Python, Chapter 6

|  |  |
| --- | --- |
| **Materials and Resources** | **Learning Goals** |
| * Computers for students with Jupyter Notebook * Slides * Exercises Jupyter Notebook file | * Understand and implement the methods listed to find the model parameters that give the best test error rate. |

|  |  |  |
| --- | --- | --- |
| **Duration** | **Lesson Section** | **Learning Objectives** |
| 20 mins | Go through the subset selection section of the slides. | * Best subset selection * Forward stepwise selection * Backward stepwise selection |
| 15 mins | Go through the best subset selection and stepwise selection sections in the Jupyter Notebook file as a class. | * Use `10bnb` package to perform best subset selection and forward stepwise selection for a linear model |
| 20 mins | Go through the indirect error estimation section of the slides. | * Indirect test error estimation (Cp, AIC, BIC, adjusted R^2) |
| 20 mins | Go through the indirection error estimation section in the Jupyter Notebook file as a class. | * Plot adjusted R2, Cp, and BIC * Interpret plots to choose the best model size. |
| 10 mins | Go through the direct error estimation section of the slides. | * Direct error estimation * Comparison of both methods |
| 25 mins | Go through the direct error estimation section in the Jupyter Notebook file as a class. | * Validation set approach estimating test error * K-folds Cross-validation for estimation test error * Choose the best model |
| 20 mins | Go through the ridge regression and the lasso sections of the slides. | * Ridge regression * The lasso * Comparison of the two * Model interpretability vs prediction accuracy * Selecting the tuning parameter |
| 30 mins | Go through the ridge regression and lasso section in the Jupyter Notebook file as a class. | * Use `skl.EleasticNet()` to perform ridge regression and the lasso * Use `GridSearchCV()` to find the best tuning parameters |