

# Exam questions:

## Chapter 7:

- How should we interpret the intercept in

$$y_i = \beta_0 + \beta_1 C_1(x_i) + \beta_2 C_2(x_i) + \dots + \beta_K C_K(x_i) + \epsilon_i.$$

When

$$\begin{aligned} C_1(X) &= I(c_1 \leq X < c_2), \\ C_2(X) &= I(c_2 \leq X < c_3), \\ &\vdots \\ C_{K-1}(X) &= I(c_{K-1} \leq X < c_K), \\ C_K(X) &= I(c_K \leq X), \end{aligned}$$

- How would you represent a cubic model in term of basis functions?
- A cubic spline has 4+K degrees of freedom. Explain!
- Do linear splines have continuous first and second derivatives?
- What is truncated power function in splines? Write down the equation for a cubic spline with 1 knot in term of the truncated power function.
- What are natural splines? What are the degrees of freedom for natural splines?
- How could you determine the optimal number of knots in spline regression?
- How would a natural spline with K=0 would look like??
- Explain the concept of a smoothing spline. Why is the penalty term needed? What happens when the penalty goes to infinity? How should you interpret the effective degrees of freedom?
- What is the relation between

$$\text{RSS}_{cv}(\lambda) = \sum_{i=1}^n (y_i - \hat{g}_{\lambda}^{(-i)}(x_i))^2 = \sum_{i=1}^n \left[ \frac{y_i - \hat{g}_{\lambda}(x_i)}{1 - \{\mathbf{S}_{\lambda}\}_{ii}} \right]^2$$

and the leverage? Give an intuitive explanation how this relation could help in computing the RSS for LOOCV?

- What have KNN and local regression in common?
- What is local regression? Explain the most influential parameter. What happens for large or small values?
- Explain the differences between the equation using basis functions:

$$y_i = \beta_0 + \beta_1 b_1(x_i) + \beta_2 b_2(x_i) + \beta_3 b_3(x_i) + \dots + \beta_K b_K(x_i) + \epsilon_i. \quad (7.7)$$

and the equation describing GAMs

$$\begin{aligned}
y_i &= \beta_0 + \sum_{j=1}^p f_j(x_{ij}) + \epsilon_i \\
&= \beta_0 + f_1(x_{i1}) + f_2(x_{i2}) + \cdots + f_p(x_{ip}) + \epsilon_i. \quad (7.15)
\end{aligned}$$

- Fitting a GAM with smoothing splines is not as simple as fitting a GAM with natural splines. Why is this the case?
- What is the main disadvantage of GAMs?
- Exercise 7.9.1, 7.9.2, 7.9.3, 7.9.4

## Chapter 8:

- What is a decision tree? Give an example how such a tree splits the predictor space? What are the branches, leaves, root, etc.? Is it possible to model interaction between predictors with a tree?
- What is recursive binary splitting? Is the scale of the predictor variables important?
- Why would we like to prune a tree? How is this performed? Why don't we just stop when the reduction in RSS drops below a pre-specified threshold?
- Explain how cost complexity pruning works? Do you compute the weakest link on the training data or test data? How is it related to Lasso? Illustrate the equation. How would you perform k-fold cross-validation?
- In which sense are classification trees and KNN similar?
- Compare regression trees and linear regression? Come up with two examples where one method would outperform the other and vice-versa.
- What is the Gini index and what is entropy? Give a formula and explain! Should we use this statistic for constructing/pruning/validating the tree or is the misclassification error preferred for some of previous steps?
- What is the purpose of bootstrap aggregation? Do we prune our trees? Do we need cross-validation when we are bagging from large datasets? Why is that?
- How could we interpret the variables of bagged trees?
- In which sense is bagging related to leave-one-out-cross-validation? They both suffer from the same unfavorable characteristic.
- What are random forests?
- What is boosting? Which hyper-parameters could you considered to tune? How do you compute out-of-bag estimates in boosted trees?