Advanced Statistical Methods
Assessment Criteria for Student Projects

General project requirements

- The projects are individual. Discussions between students concerning the projects are allowed, but any signs of code copying will lead to severe punishments.
- You have to provide a source code and a presentation with a detailed description of what steps you have done and what results you obtained. It is assumed that in the presentation you will include all the formulae and plots, which are mentioned in the criteria below. You must be able to run the code during the defense. If it does not compile, you fail the project. The source code must be clear and be accompanied by comments. The usage of iPython notebook is welcome though is not necessary.
- You do not have to implement a function if it is already implemented. If you import libraries from other languages than Python (R, C, etc.), please, provide a brief description of the imported classes/functions.
- You must follow the guidelines in the project description. If you change the way of generating the true function f^* you fail the project. The same concerns changing the sample size n or the design points.
- The project is evaluated from 10 points. Fractional points are possible. The points are not rounded.

Project 1

Student's project will be evaluated as follows:

1. (0.5 points) The data is generated correctly and according to the guidelines. The simulation is reproducible and the plot of the true function f^* and of the observations is provided.
2. (0.5 points) The family of parameters $\mathscr{M}$ is specified.
3. (1 point) The design matrix Ψ_m is computed correctly. The estimates of vector of parameters $\widetilde{\theta}_m = S_m Y$ and and of the target functional $\widetilde{\phi}_m = \mathcal{K}_m Y$ are computed correctly and explicit expressions for the matrices S_m and $\mathcal{K}_m$ are provided.

4. (0.5 points) The loss $\rho_m = \|\widetilde{\phi}_m - \phi^*\|^2$, where ϕ^* is a target functional and $\widetilde{\phi}_m$ is its estimate from the model m, is computed correctly.

5. (1 point) The risk $\mathcal{R}_m = \mathbb{E}\rho_m$ is computed correctly, an analytic expression of $\mathcal{R}_m$ and a plot indicating the dependence of $\mathcal{R}_m$ of m are provided (note that all the expectations can be computed explicitly, since you know the data distribution). The oracle choice $m^* \in \operatorname{argmin}_{m \in \mathscr{M}} \mathcal{R}_m$ is specified. Please, ensure that the minimum of $\mathcal{R}_m$ is achieved in the middle (that is, the family $\mathscr{M}$ should be chosen in a such way that m^* is not the minimal or the maximal model in $\mathscr{M}$).

6. (3 points) An adaptive choice of the parameter with the selected method is implemented correctly, a plot indicating the dependence of the risk estimate of m is provided, the selected model $\widehat{m}$ is specified. A plot with the oracle estimate and the adaptive estimate of the target (f^* of $(f^*)'$) are provided.

7. (0.5 points) The risk $\mathcal{R}_{\widehat{m}}$ is computed and compared with $\mathcal{R}_{m^*}$.

8. (1 point) The parameter choice via cross-validation is implemented correctly.

9. (2 points) Answers to the project-related questions from TA's.

Project 2

Student's project will be evaluated as follows:

1. (1 point) The data is generated correctly and according to the guidelines. The simulation is reproducible and the plot of the true function f^* and of the observations is provided. A plot with the true derivative $(f^*)'$ is provided.

2. (0.5 points) The family of parameters $\mathscr{M}$ is specified.

3. (1.5 points) The design matrix Ψ_m is computed correctly. The estimates of vector of parameters $\widetilde{\theta}_m = S_m Y$ and and of the target functional $\widetilde{\phi}_m = \mathcal{K}_m Y$ are computed correctly and explicit expressions for the matrices S_m and $\mathcal{K}_m$ are provided.

4. (0.5 points) The loss $\rho_m = \|\widetilde{\phi}_m - \phi^*\|^2$, where ϕ^* is a target functional and $\widetilde{\phi}_m$ is its estimate from the model m, is computed correctly.

5. (1 point) The risk $\mathcal{R}_m = \mathbb{E}\rho_m$ is computed correctly, an analytic expression of $\mathcal{R}_m$ and a plot indicating the dependence of $\mathcal{R}_m$ of m are provided (note that all the expectations can be computed explicitly, since you know the data distribution). The oracle choice $m^* \in \operatorname{argmin}_{m \in \mathscr{M}} \mathcal{R}_m$ is specified. Please, ensure that the minimum of $\mathcal{R}_m$ is achieved in the middle (that is, the family $\mathscr{M}$ should be chosen in a such way that m^* is not the minimal or the maximal model in $\mathscr{M}$).

6. (3 points) An adaptive choice of the parameter with the selected method is implemented correctly, a plot indicating the dependence of the risk estimate of m is provided, the selected model $\widehat{m}$ is specified. A plot with the oracle estimate and the adaptive estimate of the target (f^* of $(f^*)'$) are provided.

7. (0.5 points) The risk $\mathcal{R}_{\widehat{m}}$ is computed and compared with $\mathcal{R}_{m^*}$.

8. (2 points) Answers to the project-related questions from TA's.

Projects 3 and 4

Student's project will be evaluated as follows:

1. (0.5 points) The data is generated correctly and according to the guidelines. The simulation is reproducible and the plot of the true function f^* and of the observations is provided. In the task with derivative estimation (project 4), an additional plot with $(f^*)'$ is provided.

2. (0.5 points) The family of bandwidths $\{h_m : 1 \le m \le M\}$ and the point x_0 are specified.

3. (1 point) The estimates of vector of parameters $\widetilde{\theta}_m = S_m Y$ and and of the target functional (i.e. of $f^*(x_0)$ or of $(f^*)'(x_0)$) $\widetilde{\phi}_m = \mathcal{K}_m Y$ are computed correctly and explicit expressions for the $(1 \times n)$-matrices S_m and $\mathcal{K}_m$ are provided.

4. (0.5 points) The loss $\rho_m = (\widetilde{\phi}_m - \phi^*)^2$, where ϕ^* is a target functional ($f^*(x_0)$ or of $(f^*)'(x_0)$) and $\widetilde{\phi}_m$ is its estimate from the model m, is computed correctly.

5. (1.5 points) The risk $\mathcal{R}_m = \mathbb{E}\rho_m$ is computed correctly, an analytic expression of $\mathcal{R}_m$ and a plot indicating the dependence of $\mathcal{R}_m$ of m are provided (note that all the expectations can be computed explicitly, since you know the data distribution). The oracle choice $m^* \in \operatorname{argmin}_{m \in \mathscr{M}} \mathcal{R}_m$ is specified. Please, ensure that the minimum of $\mathcal{R}_m$ is achieved in the middle (that is, the family $\mathscr{M}$ should be chosen in a such way that m^* is not the minimal or the maximal model in $\mathscr{M}$).

6. (3 points) An adaptive choice of the parameter with the selected method is implemented correctly, a plot indicating the dependence of the risk estimate of m is provided, the selected model $\widehat{m}$ is specified.

7. (1 point) The risk $\mathcal{R}_{\widehat{m}}$ is computed and compared with $\mathcal{R}_{m^*}$.

8. (2 points) Answers to the project-related questions from TA's.

Project 5

Student's project will be evaluated as follows:

1. (0.5 points) The data is generated correctly and according to the guidelines. The simulation is reproducible and the plot of the true function f^* and of the observations is provided.

2. (0.5 points) The family of parameters $\{\tau_m : 1 \le m \le M\}$ is specified.

3. (0.5 points) The design matrix Ψ is computed correctly. The parameters of the conditional prior $(\vartheta \,|\, \tau_m)$ and of the conditional posterior distribution $(\vartheta \,|\, \tau_m, Y)$ are specified correctly.

4. (0.5 points) The hyper-posterior distribution $\exp\{L(\theta, \tau|Y)\}$ is computed correctly.

5. (1 point) A histogram of the marginal hyper-posterior $(\tau \,|\, Y)$ is provided.

6. (2 point) The frequentist risks $\mathcal{R}_m = \mathbb{E}\|\theta_m - \theta^*\|^2$, $1 \le m \le M$, are computed correctly, an analytic expression of $\mathcal{R}_m$ and a plot indicating the dependence of $\mathcal{R}_m$ of m are provided (note that all the expectations can be computed explicitly, since you know the data distribution). The oracle choice $m^* \in \operatorname{argmin}_{m \in \mathscr{M}} \mathcal{R}_m$ is specified. Please, ensure that the minimum of $\mathcal{R}_m$ is achieved in the middle (that is, the family $\mathscr{M}$ should be chosen in a such way that m^* is not the minimal or the maximal model in $\mathscr{M}$).

7. (2 points) The frequentist risk $\mathcal{R} = \mathbb{E}\|\theta - \theta^*\|^2$, where $\theta \sim (\vartheta \,|\, Y)$ is computed correctly and compared with $\mathcal{R}_{m^*}$. A plot of the function, corresponding to $\mathbb{E}(\theta \,|\, Y)$ is provided.

8. (1 point) The parameter choice via cross-validation is implemented correctly.

9. (2 points) Answers to the project-related questions from TA's.