Post-course assessment

The questions are grouped into easy and moderate/hard questions. Most questions consist of both a "theory" part and an "applied" part. Choose the questions you want to work on based on your strengths and interests. You should try at least two moderate/hard questions.

All data used in the questions on this assignment sheet can be loaded into R using the command

```
load(url("http://www.stats.gla.ac.uk/~levers/aptspca.RData"))
```

This file also contains the function `bbase` *used to construct the B-spline basis in the first practical session.*

Easy questions

Question 1 (Kernel-density estimation). Consider the distribution represented by a density estimate $\hat{f}$, constructed from a sample of data $\{y_1, \dots, y_n\}$. What is the mean and variance of this distribution? What do these expressions indicate about the nature of smoothing?

Note this question does not refer to $\mathbb{E}(\hat{f}(y))$ *and* $\text{Var}(\hat{f}(y))$ *at specific values of* y*, as discussed in the lectures. It refers to the mean and variance of a random variable whose density function is* $\hat{f}$*.*

Question 2 (Local regression). In section 4.1 (and the preliminary material) we have defined the local mean estimator as

$$\hat{m}(x) = \frac{\sum_{k=1}^n w(x_k - x; h) y_k}{\sum_{k=1}^n w(x_k - x; h)},$$

which implies that

$$\hat{y}_i = \hat{m}(x_i) = \frac{\sum_{k=1}^n w(x_k - x_i; h) y_k}{\sum_{k=1}^n w(x_k - x_i; h)}.$$

Thus the fitted values are a linear function of the observed response and we can write $\hat{\mathbf{y}} = \mathbf{S}\mathbf{y}$.

(a) Write down the entries of the matrix $\mathbf{S}$.

(b) Consider the radio-carbon dating data (available in the data frame `radiocarbon`), in which we aim to predict the age from radio-carbon dating (`rc.age`) from the calendar age (`cal.age`).

  (i) Construct the matrix $\mathbf{S}$ for a Gaussian kernel with bandwidth $h = 0.05$ and use it to compute the fitted values.

  (ii) Plot the data and add the fitted values.

  (iii) In section 3.1 of the notes we have seen that we can define the effective degrees of freedom as $\text{tr}(\mathbf{S})$. Compute the trace.

  (iv) Change the value of h. How does this change the fitted function and the effective degrees of freedom?

Question 3 (Penalised least-squares for P-splines). In section 3.3 of the notes we have introduced P-splines, which minimise the penalised least-squares criterion

$$\sum_{i=1}^n (y_i - m(x_i))^2 + \lambda \|\mathbf{D}\boldsymbol{\beta}\|^2 = \|\mathbf{y} - \mathbf{B}\boldsymbol{\beta}\|^2 + \lambda \|\mathbf{D}\boldsymbol{\beta}\|^2 = (\mathbf{y} - \mathbf{B}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{B}\boldsymbol{\beta}) + \lambda \boldsymbol{\beta}^\top \mathbf{D}^\top \mathbf{D}\boldsymbol{\beta},$$

where $\mathbf{B}$ is the matrix of B-spline basis functions, $\mathbf{D}$ is the difference matrix used in the penalty, and $\lambda \geq 0$ is the smoothing parameter.

(a) Show, by taking the derivative with respect to $\boldsymbol{\beta}$, that the minimiser of the penalised least-squares criterion is given by

$$\boldsymbol{\beta} = (\mathbf{B}^\top \mathbf{B} + \lambda \mathbf{D}^\top \mathbf{D})^{-1} \mathbf{B}^\top \mathbf{y}.$$

(b) Explain why we can rewrite the objective function as

$$\left\| \begin{pmatrix} \mathbf{B} \\ \sqrt{\lambda}\mathbf{D} \end{pmatrix} \boldsymbol{\beta} - \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} \right\|^2.$$

(c) *(harder)* How can we exploit this to estimate $\boldsymbol{\beta}$ using a QR decomposition, which is numerically more stable than inverting the matrix $\mathbf{B}^\top \mathbf{B} + \lambda \mathbf{D}^\top \mathbf{D}$?

Question 4 (Clyde dissolved oxygen data). The Clyde DO data (available in the data frame `clyde`) were used in one of the practical sessions.

Subset these data to focus only on Station 10. Fit an additive model to describe the relationship between DO and the three potential explanatory variables, Doy (day of the year), Year and Salinity. Describe the results of fitting three additive terms. Now fit a model which includes interaction (a bivariate term) between Doy and Year. Is there any evidence that the seasonal effect has changed over the years at this station?

Moderate/hard questions

Question 5 (B-spline basis functions for equally-spaced knots). The B-spline basis functions are defined recursively. Given a set of l knots the B-spline basis of degree 0 is given by the functions $(B_1^0(x), \ldots, B_{l-1}^0(x))$ with

$$B_j^0(x) = \begin{cases} 1 & \text{for } \kappa_j \leq x < \kappa_{j+1} \\ 0 & \text{otherwise.} \end{cases}$$

The B-spline basis of degree $r > 0$ is given by the functions $(B_1^r(x), \ldots, B_{l+r-1}^r(x))$ with

$$B_j^r(x) = \frac{x - \kappa_{j-r}}{\kappa_j - \kappa_{j-r}} B_{j-1}^{r-1}(x) + \frac{\kappa_{j+1} - x}{\kappa_{j+1} - \kappa_{\boldsymbol{j+1-r}}} B_j^{r-1}(x).$$

Important: Note there was a typo in the notes (subscript in bold above).

We will now turn to the important special case that the knots are equally spaced, i.e. $\kappa_{j+1} - \kappa_j = \delta$ for all j. In this case the basis functions can be computed as r-th order differences of truncated polynomials.

We start by defining the coefficients of the difference of r-th order

$$\Delta_j^r = (-1)^j \binom{r}{j} \qquad \text{for } j = 0, \ldots, r$$

For a first-order difference we obtain $\Delta_0^1 = 1$ and $\Delta_1^1 = -1$.
For a second order difference we obtain $\Delta_0^2 = 1$, $\Delta_1^2 = -2$ and $\Delta_2^2 = 1$.
For a third order difference we obtain $\Delta_0^3 = 1$, $\Delta_1^3 = -3$, $\Delta_2^3 = 3$ and $\Delta_3^3 = -1$.
These are also the numbers appearing inside the differencing matrix used in P-splines.

(a) Show that $B_j^r(x) = \sum_{i=0}^{r+1} \frac{\Delta_i^{r+1}}{r!\delta^r} (x - \kappa_{j-r+i})_+^r$

The proof is by induction using the recursive definition of B-splines.

The following two properties of Δ_j^r will help in the proof.

- $\Delta_0^{r+1} = 1$, $\Delta_{r+1}^{r+1} = (-1)^{r+1}$ *and* $\Delta_j^{r+1} = \Delta_j^r - \Delta_{j-1}^r$ *for* $j = 1, \ldots, r$.
  For first-order and second-order differences this corresponds to:

1	*-1*	
	-1	*1*
1	*-2*	*1*

  For second-order and third-order differences this corresponds to:

1	*-2*	*1*	
	-1	*2*	*-1*
1	*-3*	*3*	*-1*

  If we define $\Delta_{-1}^r = \Delta_{r+1}^r = 0$, *the formula* $\Delta_j^{r+1} = \Delta_j^r - \Delta_{j-1}^r$ *holds for all* $j = 0, \ldots, r+1$.
- $(r-i+1)\Delta_{i-1}^r = (r-i+1)(-1)^{i-1}\frac{r!}{(i-1)!(r-i+1)!} = (-1)^{i+1}\frac{r!}{(i-1)!}(r-i)! = -(-1)^i i\frac{r!}{i!(r-i)!} = -i\Delta_i^r$

(b) Write an R script (or function) which uses the above method to generate a B-spline basis of degrees 0, 1, and 2. Can you generalise your script (or function) so that it generate a B-spline basis of any order r?

Question 6 (Derivative estimation). In this question we will focus on estimating the derivative $m'(x)$ of the regression function. We will initially be focusing on B-splines.

(a) Show that

$$\frac{\partial}{\partial x}B_j^r(x) = \frac{r}{\kappa_j - \kappa_{j-r}}B_{j-1}^{r-1}(x) - \frac{r}{\kappa_{j+1} - \kappa_{j+1-r}}B_j^{r-1}(x).$$

In the special case of equally-spaced knots ($\kappa_j - \kappa_{j-1} = \delta$) this becomes

$$\frac{\partial}{\partial x}B_j^r(x) = \frac{1}{\delta}B_{j-1}^{r-1}(x) - \frac{1}{\delta}B_j^{r-1}(x).$$

Hint: The proof in the general case is done by induction using the recursive definition of B-splines. In the special case of equally spaced knot, one can simply take the derivative of the formula from question 5.

(b) Show that, in the case of equally-spaced knots,

$$\frac{\partial}{\partial x}m(x) = \sum_{j=1}^{l+r-2} B_j^{r-1}(x)\frac{\beta_{j+1} - \beta_j}{\delta},$$

i.e. we can estimate the derivative by multiplying the matrix of basis functions of degree $r-1$ with the vector $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_1, \ldots, \hat{\gamma}_{r+l-2})$, where $\hat{\gamma}_j = \dfrac{\hat{\beta}_{j+1} - \hat{\beta}_j}{\delta}$.

(c) The data frame `follicle` contains data on the number of ovarian follicles counted from sectioned ovaries of women of various ages. It has two columns.

`age`	age of the women
`log.count`	logarithm of the follicle count

  (i) Fit a B-spline model with a suitable number of knots to the `follicle` data.
  Hint: You can use the function `bbase and the code from the first practical session.`

  (ii) Suppose we are interested in the rate by which the number of follicles is reducing. We can estimate this rate by computing the derivative of the fitted regression function. Use the formula from part (b) to compute the derivative and plot it.

  (iii) Can you construct a confidence interval for the estimated derivative? From what age onward is there a significant decrease in the number of follicles

(d) Suppose you wanted to use a truncated power basis instead of a B-spline basis. How would you estimate the derivative?

(e) Suppose you wanted to use a local estimate like the one studied in question 2. How would you estimate the derivative?

Question 7 (Monotonic smoothing). Consider again the radio-carbon example in which tried to relate the observed radio-carbon age to the calibrated age. It seems natural to impose the constraint that the function describing the relationship between the two is non-decreasing. In this question you will learn how this can be achieved by using equally-spaced B-splines.

(a) Explain why the estimated regression function $\hat{m}(x) = \sum_{j=1}^{l+r-1} B_j(x)\beta_j$ is non-decreasing if $\beta_j \leq \beta_{j+1}$ for all j.
Hint: Use the derivative formula from question 6(b).

(b) This fact can be exploited to construct a monotonic regression function. "All" we need to do is to introduce the additional constraint that $\beta_j \leq \beta_{j+1}$ for all j. These additional constraints however make finding $\hat{\boldsymbol{\beta}}$ much more difficult: we have to resort to quadratic programming methods[1]. We will use a simpler approach, based on modifying the penalty in a P-splines approach.[2] When using first-order differences to construct the penalty we use

$$\mathbf{D}_1 = \begin{pmatrix} 1 & -1 & \dots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 1 & -1 \end{pmatrix},$$

i.e. the penalty becomes

$$\|\mathbf{D}_1\boldsymbol{\beta}\|^2 = \sum_{j=1}^{l+r-2} (\beta_{j+1} - \beta_j)^2.$$

In order to penalise lack of monotonicity we only want to penalise differences between the β_j if $\beta_j > \beta_{j+1}$, i.e. we would like to use the penalty

$$\sum_{j:\ \beta_j > \beta_{j+1}} (\beta_{j+1} - \beta_j)^2$$

This corresponds to modifying the matrix, setting all rows to zero that correspond to pairs with $\beta_j \leq \beta_{j+1}$.

Now there is of course the problem that we need the differencing matrix to estimate $\hat{\boldsymbol{\beta}}$, but it, in turn, depends on $\boldsymbol{\beta}$. The way around this problem is to simply iterate between these two steps, which gives the following algorithm.

1. Set $\mathbf{D}^{(1)} = \begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & 0 \\ 0 & \dots & 0 \end{pmatrix}$.

2. For $h = 1, 2, \dots$ until convergence …

   i. Compute $\boldsymbol{\beta}^{(h)} = (\mathbf{B}^\top\mathbf{B} + \lambda {\mathbf{D}^{(h)}}^\top \mathbf{D}^{(h)})^{-1}\mathbf{B}^\top\mathbf{y}$ (or use a QR decomposition to compute $\boldsymbol{\beta}^{(h)}$).

   ii. Set $\delta_j^{(h)} = \begin{cases} 1 & \text{if } \beta_j^{(h)} > \beta_{j+1}^{(h)} \\ 0 & \text{otherwise.} \end{cases}$

[1] These are for example implemented in the R package `quadprog`.

[2] This idea was first suggested by Bollaerts *et al.* (British Journal of Mathematical and Statistical Psychology (2006), 59, 451–469). Bollaerts *et al.* use a monotonicity penalty in conjunction with a smoothness penalty, but for simplicity we will omit the smoothness penalty and control the smoothness by choosing a small enough number of basis functions.

iii. Set $\mathbf{D}^{(h+1)} = \begin{pmatrix} \delta_1^{(h)} & -\delta_1^{(h)} & \dots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & \delta_{l+r-2}^{(h)} & -\delta_{l+r-2}^{(h)} \end{pmatrix}$.

We will now turn to the radiocarbon data.

(i) Fit a "classical" B-spline model with 25 equally spaced knots to the data.
Hint: You can use the function `bbase` *to create the matrix* $\mathbf{B}$.

(ii) Implement the above algorithm to estimate a non-decreasing regression function modelling the relationship between radio-carbon age and calibrated age. Use the same basis function as in part (i).

Compare the results to the B-spline model fitted in part (i).

(c) In this part we will return to the follicle data from question 6. All (primordial) follicles are developed before birth, so, after birth, the number of follicles must decrease over time.

When using a moderate number of knots (say 10) for modelling the log-follicle counts the estimated regression function is oscillating.

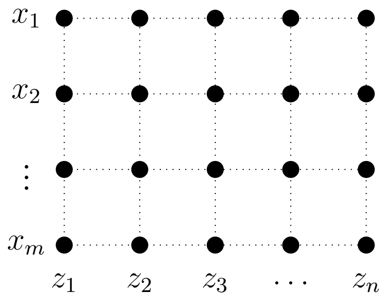
How can you use the approach set out above to estimate a *non-increasing* regression function?

Question 8 (SO_2 concentrations). The data frame `SO2station` contains measurements of SO_2 from the air at a sampling station in mainland Europe. The variables are:

`logSO2`	SO2 measurement on a log scale
`Year`	year, with a fractional component to reflect week within year
`Week`	week within the year
`Rain`	rainfall
`Temp`	temperature
`Humidity`	air humidity
`Flow`	a measurements of air flow

1. Explore the relationships between the variables, and specifically between SO_2 and potential explanatory variables, by any graphical means you consider suitable.
2. Fit an additive model and refine this into a model which you believe gives a good description of the data. (At this stage, do not worry about temporal correlation in the data.)
3. Consider a model which uses only `Year` and `Week` as explanatory variables. This model is of interest because meteorological information is not always easy to obtain, so an understanding of whether it is needed at this station may help in decisions on whether to collect it at others. Compare this reduced model with one which makes use of the meteorological information. In the comparison, interest lies particularly in any effect on the estimate of trend in SO_2 over the years.
4. Consider the additive model which contains only `Year` and `Week` as explanatory variables. Examine the residuals from this model for evidence of serial correlation. How would you adjust your model to account for this? Even if you don't do that, can you say what the effects of a suitable adjustment would be?

Question 9 (Marginal smoothing using local smoothers). Consider a bivariate smoothing problem in which the data has been collected on a regular grid, i.e. the response was observed at each combination (x_i, z_j) $(i = 1, \dots, m, j = 1, \dots, n)$. The figure below illustrates this setup.



In this case it is easiest to write the observed response as a matrix $\mathbf{Y} = (y_{ij})$ with y_{ij} being the response associated with (x_i, z_j). Alternatively we can write the observed response as a long vector, stacking the columns of $\mathbf{Y}$ on top of each other, i.e.

$$\mathbf{y} = (y_{11}, \ldots, y_{m1}, y_{12}, \ldots, y_{m2}, \ldots, y_{mn})^\top.$$

Suppose we wish to use a bivariate local smoother, i.e.

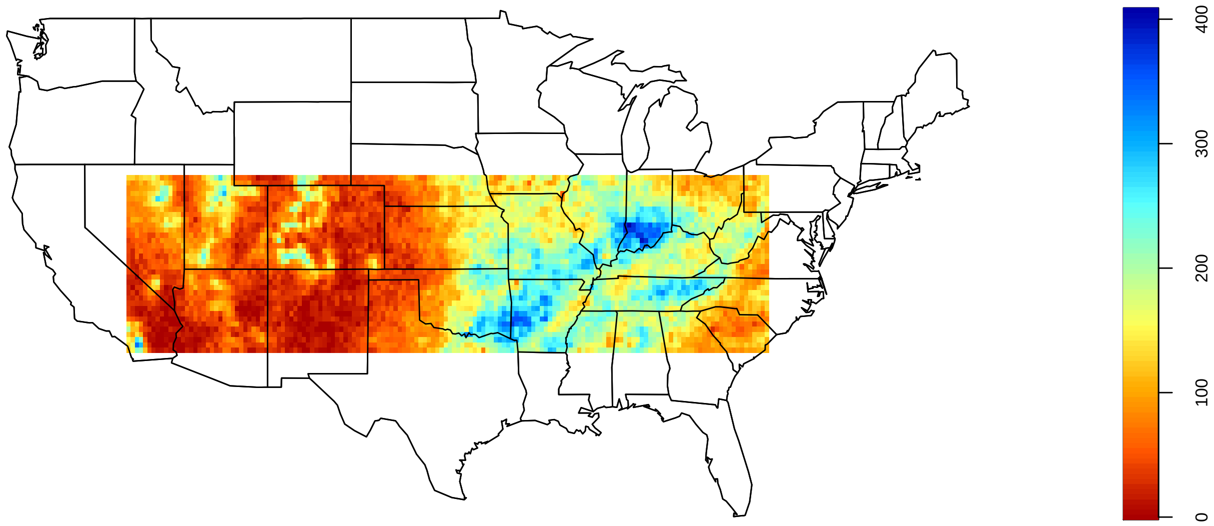
$$\hat{y}_{ij} = \hat{m}(x_i, z_j) = \frac{\sum_{k=1}^m \sum_{l=1}^n w(x_k - x_i; h) w(z_l - z_j; h) y_{kl}}{\sum_{k=1}^m \sum_{l=1}^n w(x_k - x_i; h) w(z_l - z_j; h)}.$$

(a) Just like in question 2 we can write $\hat{\mathbf{y}} = \mathbf{S}\mathbf{y}$. Write down one entry of $\mathbf{S}$.

(b) Show that $\mathbf{S} = \mathbf{S}^{(2)} \otimes \mathbf{S}^{(1)}$ where $\mathbf{S}^{(2)}$ is the smoothing matrix associated with a univariate smoothing problem with observed covariate values $z_1, \ldots, z_n$ and $\mathbf{S}^{(1)}$ is the smoothing matrix associated with a univariate smoothing problem with observed covariate values $x_1, \ldots, x_m$.
Hint: Important properties of the Kronecker product ("$\otimes$") are summarised on page 8.

(c) Using the properties of the Kronecker product show that we can also write $\hat{\mathbf{Y}} = \mathbf{S}^{(1)}\mathbf{Y}\mathbf{S}^{(2)\top}$. What is the advantage of this representation?
Hint: Think about the dimensions of the matrices involved in the calculation.

(d) The data frame `us.rain` contains noisy observations of the total rainfall in March/April 2006 in a rectangular area covering most of the central US (see figure below). The vectors `us.northing` and `us.easting` contain the corresponding latitudes and longitudes. Construct the smoothing matrices $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ and use the formula derived in part (c) to construct the fitted values. Use an R function like `image` to plot the resulting smoothed surface.



Question 10 (Marginal smoothing using B-splines and P-splines). Consider again the bivariate smoothing problem with gridded data set out in question 8. In this question we will consider the tensor-product-based spline approach, set out in section 3.4.1 of the notes.

(a) Explain that we can write the design matrix

$$\mathbf{B} = \begin{pmatrix} B_{11}(x_1, z_1) & \dots & B_{l_1+r-1,1}(x_1, z_1) & B_{12}(x_1, z_1) & \dots & B_{l_1+r-1,2}(x_1, z_1) & \dots & B_{l_1+r-1,l+2+r-1}(x_1, z_1) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ B_{11}(x_m, z_1) & \dots & B_{l_1+r-1,1}(x_m, z_1) & B_{12}(x_m, z_1) & \dots & B_{l_1+r-1,2}(x_m, z_1) & \dots & B_{l_1+r-1,l+2+r-1}(x_m, z_1) \\ B_{11}(x_1, z_2) & \dots & B_{l_1+r-1,1}(x_1, z_2) & B_{12}(x_1, z_2) & \dots & B_{l_1+r-1,2}(x_1, z_2) & \dots & B_{l_1+r-1,l+2+r-1}(x_1, z_2) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ B_{11}(x_m, z_n) & \dots & B_{l_1+r-1,1}(x_m, z_n) & B_{12}(x_m, z_n) & \dots & B_{l_1+r-1,2}(x_m, z_n) & \dots & B_{l_1+r-1,l+2+r-1}(x_m, z_n) \end{pmatrix}$$

of the bivariate tensor-product-splines as $\mathbf{B} = \mathbf{B}^{(2)} \otimes \mathbf{B}^{(1)}$, where $\mathbf{B}^{(1)}$ is a univariate B-spline basis on the x_i, $\mathbf{B}^{(2)}$ is a univariate B-spline basis on the z_j, and $B_{i,j}(x, z) = B_i^{(1)}(x) B_j^{(2)}(z)$.

(b) Suppose $\mathbf{S}$ is the bivariate smoothing matrix, i.e.

$$\mathbf{S} = \mathbf{B}(\mathbf{B}^\top \mathbf{B})^{-1}\mathbf{B}^\top$$

and $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ are the corresponding univariate smoothing matrices, i.e.

$$\mathbf{S}^{(1)} = \mathbf{B}^{(1)}(\mathbf{B}^{(1)\top}\mathbf{B}^{(1)})^{-1}\mathbf{B}^{(1)\top} \qquad \mathbf{S}^{(2)} = \mathbf{B}^{(2)}(\mathbf{B}^{(2)\top}\mathbf{B}^{(2)})^{-1}\mathbf{B}^{(2)\top}.$$

Show that $\mathbf{S} = \mathbf{S}^{(2)} \otimes \mathbf{S}^{(1)}$ and that we can write $\hat{\mathbf{y}} = \mathbf{S}\mathbf{y}$ as $\hat{\mathbf{Y}} = \mathbf{S}^{(1)}\mathbf{Y}\mathbf{S}^{(2)\top}$, just like in question 8.

(c) Is the above formula also valid if P-splines are used?

(d) Use the formula derived in part (c) to analyse the US rainfall data from question 8(d).

The Kronecker Product

Given a $m \times n$ matrix $\mathbf{C}$ and a $p \times q$ matrix $\mathbf{D}$ the *Kronecker product* of the two matrices is defined as the following $mp \times nq$ matrix:

$$\mathbf{C}\otimes\mathbf{D} = \begin{bmatrix} c_{11}\mathbf{D} & \cdots & c_{1n}\mathbf{D} \\ \vdots & \ddots & \vdots \\ c_{m1}\mathbf{D} & \cdots & c_{mn}\mathbf{D} \end{bmatrix} = \begin{bmatrix} c_{11}d_{11} & c_{11}d_{12} & \cdots & c_{11}d_{1q} & \cdots & c_{1n}d_{11} & c_{1n}d_{12} & \cdots & c_{1n}d_{1q} \\ c_{11}d_{21} & c_{11}d_{22} & \cdots & c_{11}d_{2q} & \cdots & c_{1n}d_{21} & c_{1n}d_{22} & \cdots & c_{1n}d_{2q} \\ \vdots & \vdots & \ddots & \vdots & & & \vdots & \vdots & \ddots & \vdots \\ c_{11}d_{p1} & c_{11}d_{p2} & \cdots & c_{11}d_{pq} & \cdots & c_{1n}d_{p1} & c_{1n}d_{p2} & \cdots & c_{1n}d_{pq} \\ \vdots & \vdots & & \vdots & \ddots & \vdots & \vdots & & \vdots \\ c_{m1}d_{11} & c_{m1}d_{12} & \cdots & c_{m1}d_{1q} & \cdots & c_{mn}d_{11} & c_{mn}d_{12} & \cdots & c_{mn}d_{1q} \\ c_{m1}d_{21} & c_{m1}d_{22} & \cdots & c_{m1}d_{2q} & \cdots & c_{mn}d_{21} & c_{mn}d_{22} & \cdots & c_{mn}d_{2q} \\ \vdots & \vdots & \ddots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ c_{m1}d_{p1} & c_{m1}d_{p2} & \cdots & c_{m1}d_{pq} & \cdots & c_{mn}d_{p1} & c_{mn}d_{p2} & \cdots & c_{mn}d_{pq} \end{bmatrix}$$

For this assignment you will need the following properties of the Kronecker product.

- The Kronecker product is, just like the standard matrix product, *not* commutative, i.e. $\mathbf{C}\otimes\mathbf{D} \neq \mathbf{D}\otimes\mathbf{C}$. It is however associative, i.e. $(\mathbf{C}\otimes\mathbf{D})\otimes\mathbf{E} = \mathbf{C}\otimes(\mathbf{D}\otimes\mathbf{E})$.
- $\mathbf{C}\otimes(\mathbf{D}+\mathbf{E}) = \mathbf{C}\otimes\mathbf{D} + \mathbf{C}\otimes\mathbf{E}$ and $(\mathbf{C}+\mathbf{D})\otimes\mathbf{E} = \mathbf{C}\otimes\mathbf{E} + \mathbf{D}\otimes\mathbf{E}$.
- $(\mathbf{C}\otimes\mathbf{D})(\mathbf{E}\otimes\mathbf{F}) = (\mathbf{CE})\otimes(\mathbf{DF})$.
- $(\mathbf{C}\otimes\mathbf{D})^{-1} = \mathbf{C}^{-1}\otimes\mathbf{D}^{-1}$.
- $(\mathbf{C}\otimes\mathbf{D})^{\top} = \mathbf{C}^{\top}\otimes\mathbf{D}^{\top}$.
- $(\mathbf{A}\otimes\mathbf{B})\mathbf{c} = \mathbf{d}$ if and only if $\mathbf{BCA}^{\top} = \mathbf{D}$, where the vectors $\mathbf{c}$ (and $\mathbf{d}$) simply consist of the columns of $\mathbf{C}$ (and $\mathbf{D}$) stacked on top of each other, i.e.

  $$\mathbf{C} = \begin{pmatrix} c_{11} & \dots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{m1} & \dots & c_{mn} \end{pmatrix} \qquad \mathbf{c} = (c_{11}, \dots, c_{m1}, c_{12}, \dots, c_{m2}, \dots, c_{mn})^{\top}$$

  $$\mathbf{D} = \begin{pmatrix} d_{11} & \dots & d_{1q} \\ \vdots & \ddots & \vdots \\ d_{p1} & \dots & d_{pq} \end{pmatrix} \qquad \mathbf{d} = (d_{11}, \dots, d_{p1}, d_{12}, \dots, d_{p2}, \dots, d_{pq})^{\top}$$

  $\mathbf{A}$ is a $q \times n$ matrix and $\mathbf{B}$ is a $p \times m$ matrix, thus $\mathbf{C}$ is a $m \times n$ matrix and $\mathbf{D}$ is a $p \times q$ matrix.

  In R this can be verified as follows.

```r
# Create example matrices
A <- matrix(rnorm(6), ncol=3)
B <- matrix(rnorm(8), ncol=2)
C <- matrix(rnorm(6), ncol=3)
c <- as.vector(C)
# Calculate d
kronecker(A,B)%*%c
# Calculate D
B%*%C%*%t(A)
```