Correction Tutorial 2

1. See the R code.
2. See the R code.
3. We are in fact using the model
$$Y = m + \sigma\varepsilon$$
with $\varepsilon \sim \mathcal{N}_n(0, I_n)$ and $m \in V = \text{vect}\{e_1, ..., e_{2d+1}\}$, where for $k = 1, ..., d$,
$$e_1 = \frac{1}{\sqrt{n}}\begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix}, \; e_{2k} = \frac{1}{\sqrt{n/2}}\begin{pmatrix}\cos(ku_0)\\\cos(ku_1\\\vdots\\\cos(ku_{n-1})\end{pmatrix}, \; e_{2k+1} = \frac{1}{\sqrt{n/2}}\begin{pmatrix}\sin(ku_0)\\\sin(ku_1)\\\vdots\\\sin(ku_{n-1})\end{pmatrix}$$
that corresponds to the first $2d + 1$ columns of X'. Since they are prthonormal, we know that the projection estimator is given by
$$\Pi_V Y = \langle Y, e_1\rangle e_1 + ... + \langle Y, e_{2d+1}\rangle e_{2d+1}$$
4. See the R code. Clearly when $d = 40$, we get a very noisy estimator. This estimator tries to explains the fluctuations in the data that are in fact due to noise : this is the overfitting phenomenon. When d is too small (especially in the second example with $d = 1$), clearly, we do not have enough variability to explain the data. That is why we have to do the good compromise between bias and variance.
5. See the R code.
6. We are interested in the Gaussian model
$$Y_i = \underbrace{a_1 + a_2\cos(u_i) + a_2\sin(u_i) + ... + a_{2p}\cos(pu_i) + a_{2p+1}\sin(pu_i)}_{m_i} + \sigma\varepsilon_i$$
with $m \in V_p := \text{vect}\{e_1, ..., e_{2p+1}\}$. The log-likelihood can be written
$$\ell(m) = -\frac{\|Y - m\|^2}{2\sigma^2} - \frac{1}{2}\log(\sigma^2) - \frac{1}{2}\log(2\pi)$$
but if instead of V_p we consider a subset of indices $\Omega = \{2, 3, 5, 6\}$ or whatever the set included in $\{1, ..., 2p + 1\}$, then we can look at the subspace $V_r = \text{vect}\{e_i, i \in \Omega\}$, and the log likelihood will be the same. The Bic criterion then consists in looking at
$$-\max_{m\in V_r}\ell(m) + \frac{\#r}{2}\log(n)$$

where $\#r$ is the number of parameters. We have

$$
\begin{aligned}
-\max_{m\in V_r}\ell(m)+\frac{\#r}{2}\log(n) &= \frac{\|Y-\Pi_{V_r}Y\|^2}{2\sigma^2}+\frac{1}{2}\log(\sigma^2)+\frac{1}{2}\log(2\pi)+\frac{\#r}{2}\log(n)\\
&= \sum_{\substack{i=1\\ i\notin r}}^{2p+1}\frac{\langle Y,e_i\rangle^2}{2\sigma^2}+\frac{1}{2}\log(\sigma^2)+\frac{1}{2}\log(2\pi)+\frac{\#r}{2}\log(n)\\
&= \sum_{i=1}^{2p+1}\frac{\langle Y,e_i\rangle^2}{2\sigma^2}+\frac{1}{2}\log(\sigma^2)+\frac{1}{2}\log(2\pi)+\sum_{i\in r}\left[\frac{1}{2}\log(n)-\frac{\langle Y,e_i\rangle^2}{2\sigma^2}\right].
\end{aligned}
$$

If we want to minimize this quantity in r, we have to find the largest collection r such that the last sum here above is as negative as possible. For that purpose, we take eah i such that

$$
\frac{\langle Y,e_i\rangle^2}{2\sigma^2}>\frac{1}{2}\log(n),
$$

i.e

$$
\left|\langle Y,e_i\rangle\right|>\sqrt{\sigma^2\log(n)}.
$$

So the estimator becomes

$$
\sum_{i=1}^{2p+1}\left[\langle Y,e_i\rangle\mathbb{1}_{\left\{\left|\langle Y,e_i\rangle\right|>\sqrt{\sigma^2\log(n)}\right\}}\right]e_i.
$$

N.B In fact it is as if we were doing a sort of "coordinate" per coordinate test as in the usual R command `lm()`. The "$\log(n)$" can be interpreted as a Bonferroni correction.

Remark It is a bit too noisy. We can add a parameter γ and choose it by cross validation, but I prefer to show you that on another exercise.