EE 263 Homework 3

Luis A. Perez

Fitting a model for hourly temperature

Solution:

(a) In order to find $a \in \mathbb{R}$ and $p \in \mathbb{R}^N$ (which is 24-periodic) that minimizes the RMS value of $y - \hat{y}$, we can rephrase our original predictor model as a linear system:

$$\hat{y} = Ax$$

where $\hat{y} \in \mathbb{R}^N$ and $x \in \mathbb{R}^{25}$, which represents our parameters (since p is 24-periodic). More precisely, we have:

$$x = \begin{bmatrix} p_{24} \\ p_{23} \\ \vdots \\ p_2 \\ p_1 \\ a \end{bmatrix} \in \mathbb{R}^25$$

What is the same of A. In fact, we have the following:

$$A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & N \\ 0 & 1 & 0 & \cdots & 0 & N-1 \\ 0 & 0 & 1 & \cdots & 0 & N-2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & N-23 \\ 1 & 0 & 0 & \cdots & 0 & N-24 \\ 0 & 1 & 0 & \cdots & 0 & N-25 \\ 0 & 0 & 1 & \cdots & 0 & N-26 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \end{bmatrix} \in \mathbb{R}^{N \times 25}$$

This means that A is a skinny and tall matrix, for an over-constrained system of equations. Finding the x that minimizes the RMS of $\hat{y} - y$ can be done by computing:

$$x = (A^T A)^{-1} A^T y$$

The x above give use p as well as a.

(b) We now perform the process described in part (a). The trend parameter is:

$$a = -0.012075460503471858$$

A plot of the predictions as well as the observed values can be seen in Figure 1.

(c) The RMSE of the prediction error for tomorrow's temperatures is:

$$0.6521628280735887$$

A plot of the predicted and observed temperatures for tomorrow can be seen in Figure 2.

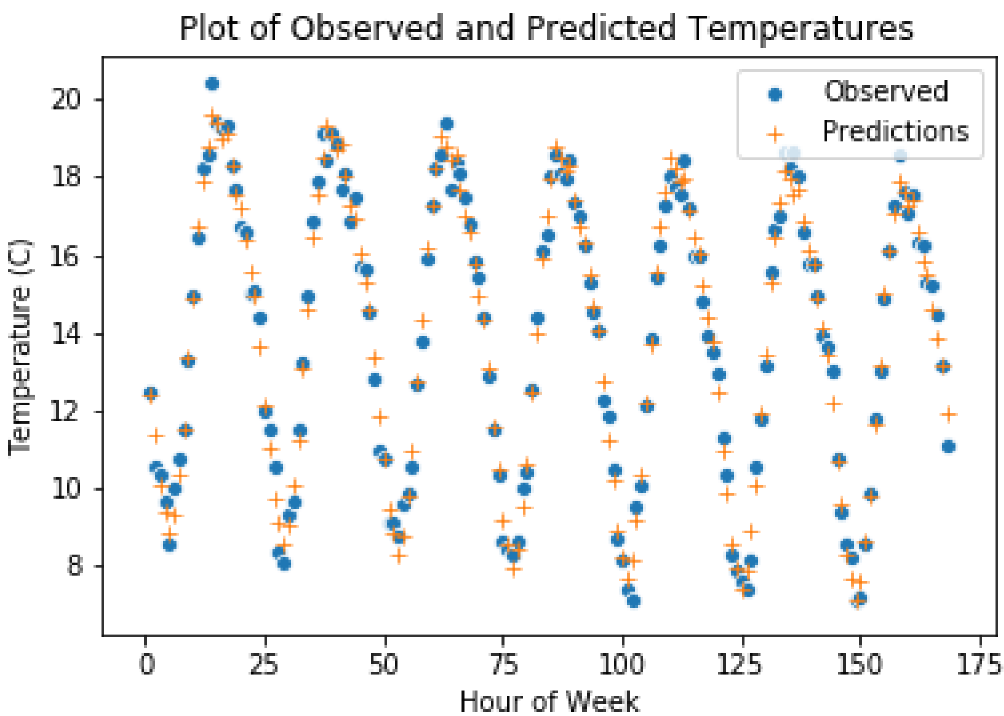

Figure 1: Plot of predicted (x) and observed (o) temperatures on training data.

Identifying a system from input/output data

Solution:

(a) We're asked to estimate the matrix A, which is somewhat non-standard (we have tools readily available for finding x, but finding A isn't quite as straight-forward).

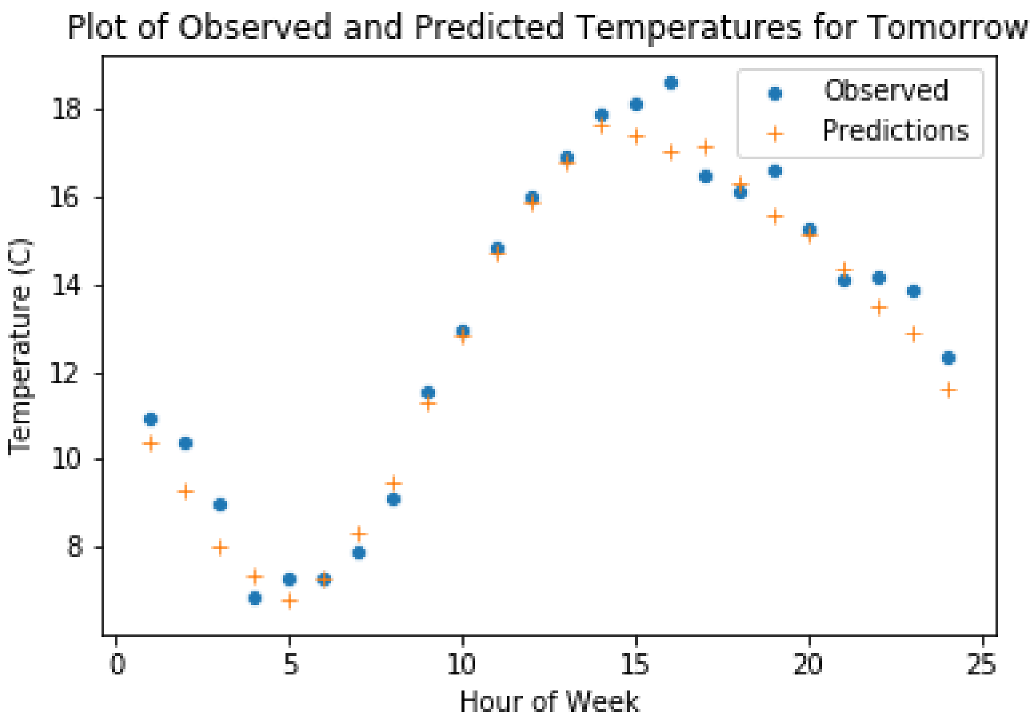

Figure 2: Plot of predicted (x) and observed (o) temperatures on test data.

We beging by rewriting the objective in a slightly different way:

$$\begin{aligned}
J &= \sum_{k=1}^{N} ||Ax^{(k)} - y^{(k)}||^2 \\
&= \sum_{k=1}^{N} \sum_{i=1}^{m} (Ax^{(k)} - y^{(k)})_i^2 && \text{(Definition of norm)} \\
&= \sum_{k=1}^{N} \sum_{i=1}^{m} (a_i^T x^{(k)} - y^{(k)})_i^2 && \text{(Where } a_i^T \text{ is the } i\text{-th row of } A) \\
&= \sum_{i=1}^{m} \left(\sum_{k=1}^{N} (a_i^T x^{(k)} - y^{(k)})_i^2 \right) && \text{(Swapping summations)} \\
&= \sum_{i=1}^{m} J_i
\end{aligned}$$

From the above, we can now see that J (our objective) is really just the sum of multiple (m) independent objectives, which we label J_i above. Since each J_i is independent of the other, minimizing J consists simply of minimizing each of the J_i.

However, note that that we can actually express J_i in the matrix form:

$$J_i = ||Xa_i - y_i||^2$$

where :

$$y_i = \begin{bmatrix} y_i^{(1)} \\ y_i^{(2)} \\ \vdots \\ y_i^{(N)} \end{bmatrix} \in \mathbb{R}^N$$

$$X = \begin{bmatrix} \cdots & x^{(1)} & \cdots \\ \cdots & x^{(2)} & \cdots \\ & \vdots & \\ \cdots & x^{(N)} & \cdots \end{bmatrix} \in \mathbb{R}^{N \times n}$$

It should be almost immediate from the above format that finding a_i to minimize J_i is just finding the least-squares solution to this over-constrained system. We can do this with the formula:

$$\hat{a}_i = (X^T X)^{-1} X^T y_i$$

(b) TODO – bad method.

Robust regression using the Huber penalty function

Solution:

(a) TODO

Identifying a system from input/output data

Solution:

(a) TODO