ML Ass1

Kuan Teh Wan

April 2024

1 Exercise 1

By differentiating

$$\underset{\omega_i}{\text{minimize}}\ \frac{1}{2}\|r_i - x_i\omega_i\|_2^2 + \lambda|\omega_i|$$

w.r.t ω_i, we have that

$$\begin{aligned}
&\frac{d}{d\omega_i}\left(\frac{1}{2}\|r_i - x_i\omega_i\|_2^2 + \lambda|\omega_i|\right) \\
&= \frac{1}{2}(2(r_i - x_i\omega_i)(-x_i) + \lambda\frac{\omega_i}{|\omega_i|}) \\
&= (r_i - x_i\omega_i)(-x_i) + \lambda\frac{\omega_i}{|\omega_i|} \\
&= -x_i^T(r_i - x_i\omega_i) + \lambda\frac{\omega_i}{|\omega_i|}
\end{aligned} \tag{1}$$

Then if (1) equals to 0, we will have the minimum we're looking for. Thus, according to (1)

$$\begin{aligned}
-x_i^T(r_i - x_i\omega_i) + \lambda\frac{\omega_i}{|\omega_i|} &= 0 \\
x_i^T r_i &= \omega_i(x_i^T x_i + \lambda\frac{1}{|\omega_i|})
\end{aligned} \tag{2}$$

Consider both sides of the equation in their absolute value and since $x_i^T x_i$ and λ are already positive, we have

$$\begin{aligned}
|x_i^T r_i| &= |\omega_i|(x_i^T x_i + \lambda\frac{1}{|\omega_i|}) \\
|\omega_i| &= \frac{|x_i^T r_i| - \lambda}{x_i^T x_i}
\end{aligned} \tag{3}$$

From the last line of (2) we can derive $sgn(\omega_i)$ since the parentheses contains only positive values.

$$sgn(\omega_i) = sgn(x_i^T r_i) = \frac{x_i^T r_i}{|x_i^T r_i|} \tag{4}$$

We then have the optimal ω_i by multiplying (3) and (4)

$$\omega_i = \frac{x_i^T r_i}{x_i^T x_i |x_i^T r_i|}(|x_i^T r_i| - \lambda) \tag{5}$$

which is what we wanted to verify.

2 Exercise 2

We may rewrite the equation for $|x_i^T r_i^{(j)}| > \lambda$

$$\begin{aligned}\hat{\omega_i}^{(j)} &= \frac{x_i^T r_i^{(j-1)}}{x_i^T x_i |x_i^T r_i^{(j-1)}|}(|x_i^T r_i^{(j-1)}| - \lambda) \\ &= x_i^T r_i^{(j-1)} - \lambda sgn(x_i^T r_i^{(j-1)})\end{aligned} \tag{6}$$

And with the information given in the assignment that one can rewrite

$$x_i^T r_i^{(j-1)} = x_i^T (t - \sum_{\ell < i} x_\ell \hat{\omega}_\ell^{(j)} - \sum_{\ell > i} x_\ell \hat{\omega}_\ell^{(j-1)}) = x_i^T t \tag{7}$$

This stands because that $x_i^T x_l = 0 \ \forall l \neq i$ and $x_i^T x_i = 1$. Thus we can further simply (6) to

$$\hat{\omega_i}^{(j)} = x_i^T t - \lambda sgn(x_i^T t) \tag{8}$$

Therefore, $\hat{\omega_i}^{(j)}$does not depend on previous estimates since

$$\hat{\omega_i}^{(2)} - \hat{\omega_i}^{(1)} = x_i^T t - \lambda sgn(x_i^T t) - (x_i^T t - \lambda sgn(x_i^T t)) = 0 \tag{9}$$

but only depends on t, x_i, and λ. And for the case that $|x_i^T r_i^{(j)}| < \lambda$, $\hat{\omega_i}^{(j)} = 0$.

3 Exercise 3

In this exercise, we would like to show that

$$\lim_{\sigma \to 0} E\left(\hat{\omega}_i^{(1)} - \omega_i^*\right) = \begin{cases} -\lambda, & \omega_i^* > \lambda \\ -\omega_i^*, & |\omega_i^*| \leq \lambda \\ \lambda, & \omega_i^* < -\lambda \end{cases} \quad \forall i \tag{10}$$

As the hint suggested, we start by investigating the first 2 cases in the closed-form solution then further break it down into 3 cases.

3.1 Case 1: $x_i^T r_i^{(j-1)} > \lambda$

From (8) we have that

$$\lim_{\sigma \to 0} x_i^T = x_i^T X\omega^* = \omega_i^* \tag{11}$$

The expected value can be obtained

$$\begin{aligned} &\lim_{\sigma \to 0} E\left(\hat{\omega}_i^{(1)} - \omega_i^*\right) \\ &= \lim_{\sigma \to 0} E\left(x_i^T t - \lambda sgn(x_i^T t) - \omega_i^*\right) \\ &= \lim_{\sigma \to 0} E\left(\omega_i^* - \lambda sgn(\omega_i^*) - \omega_i^*\right) \\ &= -\lambda sgn(\omega_i^*) = -\lambda \end{aligned} \tag{12}$$

And $x_i^T r_i^{(j-1)} \to \omega_i^*$ when $\sigma \to 0$ so the last line stands.

3.2 Case 2: $x_i^T r_i^{(j-1)} < \lambda$

This case we do the same derivation as Case 1, but with a changed sign of ω_i

$$\begin{aligned} &\lim_{\sigma \to 0} E\left(\hat{\omega}_i^{(1)} - \omega_i^*\right) \\ &= \lim_{\sigma \to 0} E\left(x_i^T t - \lambda sgn(x_i^T t) - \omega_i^*\right) \\ &= \lim_{\sigma \to 0} E\left(\omega_i^* - \lambda sgn(\omega_i^*) - \omega_i^*\right) \\ &= -\lambda sgn(\omega_i^*) = \lambda \end{aligned} \tag{13}$$

3.3 Case 3: $|x_i^T r_i^{(j-1)}| \leq \lambda$

For this case, we take $\hat{\omega_i}^{(1)}$=0. Thus,

$$\lim_{\sigma \to 0} E\left(\hat{\omega}_i^{(1)} - \omega_i^*\right) = \lim_{\sigma \to 0} E\left(0 - \omega_i^*\right) = -\omega_i^* \tag{14}$$

which fulfills equation (10). We can see that as the regularization term λ increases, the estimate's bias will also increase. Therefore, be must aware that there's a trade-off relationship between bias and variance when using the LASSO model.

4 Exercise 4

The original data is sampled from a sum of two complex-valued sinusoids. We're reconstructing the data using LASSO regularization with different lambda values. Plotted using $\lambda = 0.1$, 10 and finally 1.5 to compare.

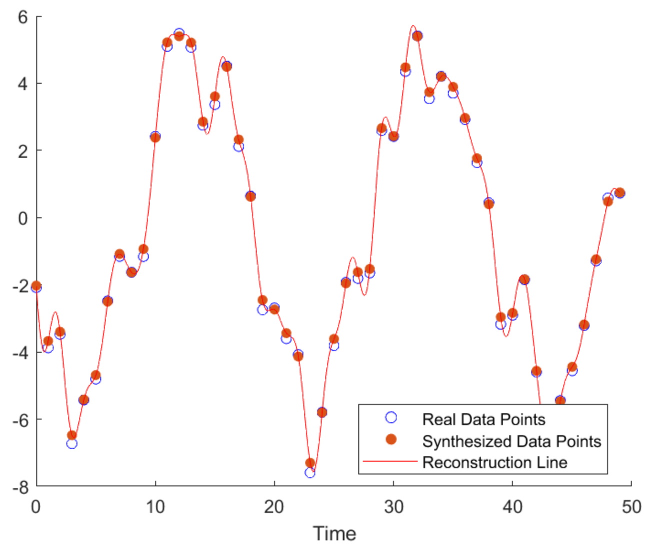

Figure 1: Reconstruction with $\lambda = 0.1$

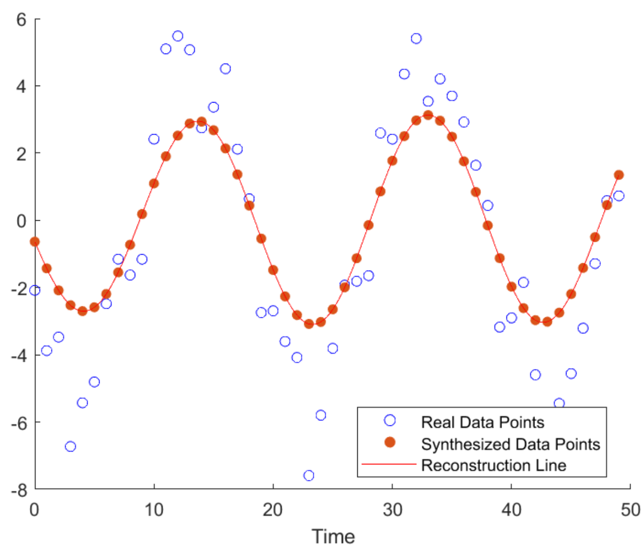

Figure 2: Reconstruction with $\lambda = 10$

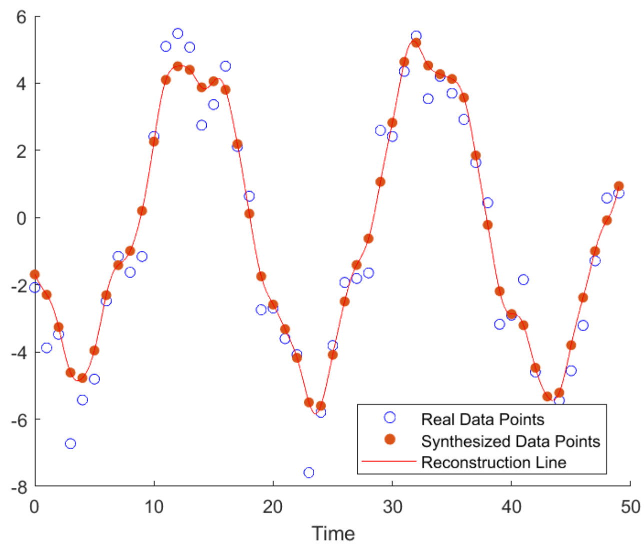

Figure 3: Reconstruction with $\lambda = 1.5$

From figure 1, we clearly see that the model here is overfitting as the synthesized data are almost overlapping with the original one, which means it is adapting to the noise instead of learning from the real data.

On the other hand, figure 2 here, shows clear sign of an underfitting model as the synthesized data are barely touching the real data.

Then, in figure 3, seems to be a more adequate model to capture the signal since it is relatively close to the real data yet not overlapping.

λ	no. of non-zero
0.1	233
10	9
1.5	33

From the table above, we can see that the larger the λ value, is the more non-zero coordinates we'll need for the model. In conclusion, we'll need more than than the 4 coordinates to have a more reasonable reconstruction.

5 Exercise 5

To find the optimal λ for the reconstruction, the $RMSE_{val}$ has to be minimized. This is archived by using 10-fold cross validation with 200 different λ values, ranging from $[0.01, max(|X^T t|)]$.Figures are shown below.

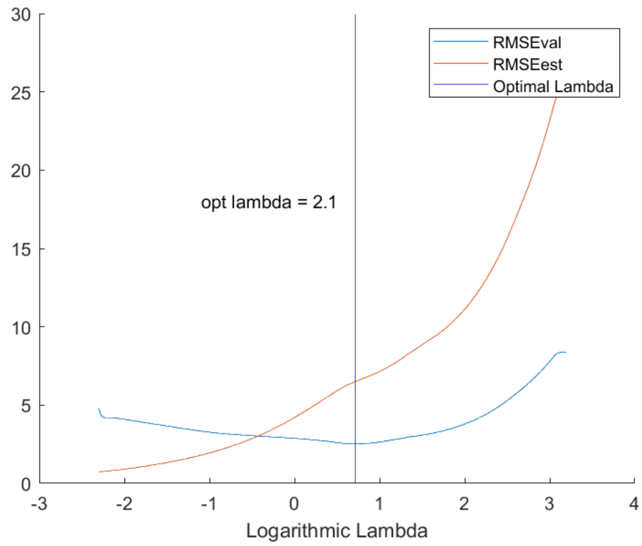

Figure 4: RMSE of Validation and Estimation using 10-fold cross validation different lambda values.(log scale)

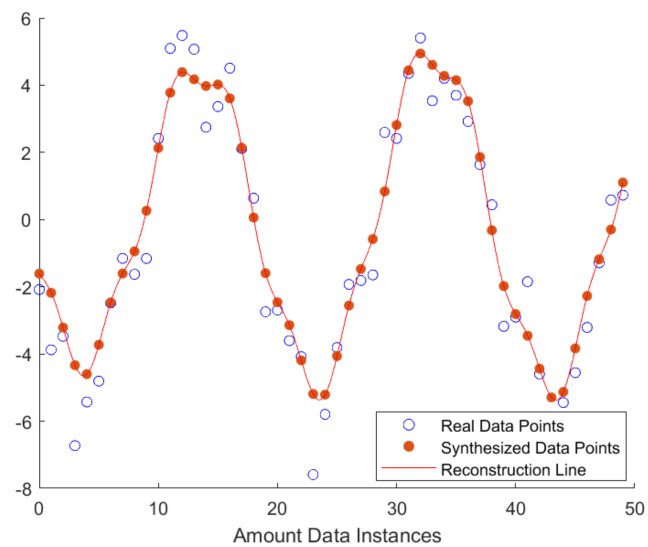

Figure 5: Reconstruction using optimal $\lambda = 2.1$

6 Exercise 6

Using multiframe 3-fold validation with 100 different λ values, ranging from $[0.001, max(|X^T t_i|)]$, where i is the number of frames, then it is possible to calculate $RMSE_{val}$ and $RMSE_{est}$ for all λ values.

To find the optimal λ, simply locate where the lowest $RMSE_{val}$ occurs, in this case, the optimal λ will be 0.0041.

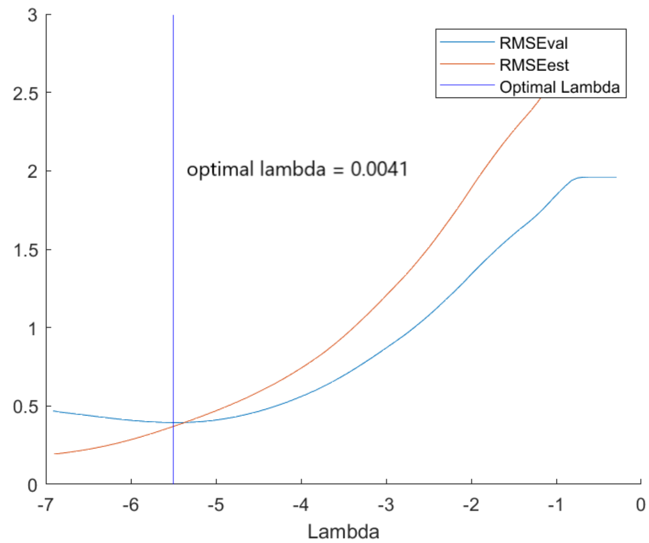

Figure 6 : RMSE of Validation and Estimation of different lambda values(log scale).

As shown in the figure above, can see that no or little regularization leads to a higher level of $RMSE_{val}$ so as strong regularization. Choosing an adequate regularization is need to have a clear reconstruction of the audio.

7 Exercise 7

After listening to the denoised version of the audio, compared to the original one, the noise is filtered out more and can hear the piano much more clearer. And by adjusting the lambda values up and down a little, both make the audio more distorted, which means the optimal we found is pretty much the best in this case.