02471 Machine Learning for Signal Processing

Solution

Exercise 5: Adaptive Linear Filtering with RLS

5.1 Derivation of the RLS algorithm

Exercise 5.1.1

For $\beta = 1$.

Exercise 5.1.2

If $0 < \beta < 1$, the regularization term decreases as we observe more data. The motivation is that for large amounts of data, the risk of overfitting is reduced, so the regularization term is not required.

Exercise 5.1.3

The cost function is defined as:

$$J(\boldsymbol{\theta}, \beta, \lambda) = \sum_{i=0}^{n} \beta^{n-i}(y_i - \boldsymbol{\theta}^T \boldsymbol{x}_i)^2 + \lambda \beta^{n+1} ||\boldsymbol{\theta}||^2$$

Taking the derivative with respect to $\boldsymbol{\theta}$ gives:

$$\begin{aligned}
\frac{\partial J(\boldsymbol{\theta}, \beta, \lambda)}{\partial \boldsymbol{\theta}} &= \sum_{i=0}^{n} \beta^{n-i} \frac{\partial}{\partial \boldsymbol{\theta}} \Big((y_i - \boldsymbol{\theta}^T \boldsymbol{x}_i)^2 \Big) + \lambda \beta^{n+1} \frac{\partial}{\partial \boldsymbol{\theta}} \big(||\boldsymbol{\theta}||^2 \big) \\
&= -2 \sum_{i=0}^{n} \beta^{n-i} (y_i - \boldsymbol{\theta}^T \boldsymbol{x}_i) \boldsymbol{x}_i + 2\lambda \beta^{n+1} \boldsymbol{\theta} \\
&= -2 \sum_{i=0}^{n} \beta^{n-i} y_i \boldsymbol{x}_i + 2 \sum_{i=0}^{n} \beta^{n-i} \underbrace{(\boldsymbol{\theta}^T \boldsymbol{x}_i) \boldsymbol{x}_i}_{\substack{= \boldsymbol{x}_i (\boldsymbol{\theta}^T \boldsymbol{x}_i) \\ = \boldsymbol{x}_i (\boldsymbol{x}_i^T \boldsymbol{\theta}) \\ = (\boldsymbol{x}_i \boldsymbol{x}_i^T) \boldsymbol{\theta}}} + 2\lambda \beta^{n+1} \boldsymbol{\theta} \\
&= \underbrace{-2 \sum_{i=0}^{n} \beta^{n-i} y_i \boldsymbol{x}_i}_{= \boldsymbol{p}_n} + 2 \underbrace{\left(\sum_{i=0}^{n} \beta^{n-i} \boldsymbol{x}_i \boldsymbol{x}_i^T + \lambda \beta^{n+1} I \right)}_{= \Phi_n} \boldsymbol{\theta}
\end{aligned}$$

Setting the derivative to 0 on $\boldsymbol{\theta}_n$ gives:

$$\Phi_n \boldsymbol{\theta}_n = \boldsymbol{p}_n$$

Exercise 5.1.4

Let's start from the formula for Φ_n and make Φ_{n-1} appear:

$$\begin{aligned}
\Phi_n &= \sum_{i=0}^{n} \beta^{n-i}\boldsymbol{x}_i\boldsymbol{x}_i^T + \lambda\beta^{n+1}I \\
&= \sum_{i=0}^{n-1} \beta^{n-i}\boldsymbol{x}_i\boldsymbol{x}_i^T + \beta^{n-n}\boldsymbol{x}_n\boldsymbol{x}_n^T + \lambda\beta^{n+1}I \\
&= \sum_{i=0}^{n-1} \beta\beta^{n-1-i}\boldsymbol{x}_i\boldsymbol{x}_i^T + \boldsymbol{x}_n\boldsymbol{x}_n^T + \lambda\beta\beta^{n}I \\
&= \beta\left(\sum_{i=0}^{n-1} \beta^{n-1-i}\boldsymbol{x}_i\boldsymbol{x}_i^T + \lambda\beta^{n}I\right) + \boldsymbol{x}_n\boldsymbol{x}_n^T \\
&= \beta\Phi_{n-1} + \boldsymbol{x}_n\boldsymbol{x}_n^T
\end{aligned}$$

Similarly for $\boldsymbol{p}_n$:

$$\begin{aligned}
\boldsymbol{p}_n &= \sum_{i=0}^{n} \beta^{n-i}y_i\boldsymbol{x}_i \\
&= \sum_{i=0}^{n-1} \beta^{n-i}y_i\boldsymbol{x}_i + \beta^{n-n}y_n\boldsymbol{x}_n \\
&= \beta\sum_{i=0}^{n-1} \beta^{n-1-i}y_i\boldsymbol{x}_i + y_n\boldsymbol{x}_n \\
&= \beta\boldsymbol{p}_{n-1} + y_n\boldsymbol{x}_n
\end{aligned}$$

Exercise 5.1.5

From Appendix 1, we have:

$$(A + BD^{-1}C)^{-1} = A^{-1} - A^{-1}B(D + CA^{-1}B)^{-1}CA^{-1}$$

Let's use $A = \beta\Phi_{n-1}$, $B = \boldsymbol{x}_n$, $C = \boldsymbol{x}_n^T$ and $D = 1$. The left term becomes:

$$(A + BD^{-1}C)^{-1} = (\beta\Phi_{n-1} + \boldsymbol{x}_n\boldsymbol{x}_n^T)^{-1}$$

which is exactly Φ_n^{-1} according to the previous question. As for the right term, it becomes:

$$\begin{aligned}
A^{-1} - A^{-1}B(D + CA^{-1}B)^{-1}CA^{-1} &= \beta^{-1}\Phi_{n-1}^{-1} - \beta^{-1}\Phi_{n-1}^{-1}\boldsymbol{x}_n\left(1 + \boldsymbol{x}_n^T\beta^{-1}\Phi_{n-1}^{-1}\boldsymbol{x}_n\right)^{-1}\boldsymbol{x}_n^T\beta^{-1}\Phi_{n-1}^{-1} \\
&= \beta^{-1}\Phi_{n-1}^{-1} - \beta^{-1}\frac{\beta^{-1}\Phi_{n-1}^{-1}\boldsymbol{x}_n}{1 + \boldsymbol{x}_n^T\beta^{-1}\Phi_{n-1}^{-1}\boldsymbol{x}_n}\boldsymbol{x}_n^T\Phi_{n-1}^{-1} \\
&= \beta^{-1}\Phi_{n-1}^{-1} - \beta^{-1}\boldsymbol{k}_n\boldsymbol{x}_n^T\Phi_{n-1}^{-1}
\end{aligned}$$

where we introduced $\boldsymbol{k}_n = \dfrac{\beta^{-1}\Phi_{n-1}^{-1}\boldsymbol{x}_n}{1 + \beta^{-1}\boldsymbol{x}_n^T\Phi_{n-1}^{-1}\boldsymbol{x}_n}$.

Equating the left and right terms thus gives the desired result:

$$\Phi_n^{-1} = \beta^{-1}\Phi_{n-1}^{-1} - \beta^{-1}\boldsymbol{k}_n\boldsymbol{x}_n^T\Phi_{n-1}^{-1}$$

Exercise 5.1.6

From 5.1.3 we have:

$$\boldsymbol{\theta}_n = \Phi_n^{-1}\boldsymbol{p}_n$$

Plugging the recursive formula for $\boldsymbol{p}_n$ found in 5.1.4 gives:

$$\begin{aligned}\boldsymbol{\theta}_n &= \Phi_n^{-1}\left(\beta\boldsymbol{p}_{n-1} + \boldsymbol{x}_n y_n\right) \\ &= \beta\Phi_n^{-1}\boldsymbol{p}_{n-1} + \Phi_n^{-1}\boldsymbol{x}_n y_n\end{aligned}$$

Let's now plug the expression for Φ_n^{-1} found in the previous question, only for the left term of the sum for now. This makes $\boldsymbol{\theta}_{n-1}$ appear:

$$\begin{aligned}\boldsymbol{\theta}_n &= \beta\left(\beta^{-1}\Phi_{n-1}^{-1} - \beta^{-1}\boldsymbol{k}_n\boldsymbol{x}_n^T\Phi_{n-1}^{-1}\right)\boldsymbol{p}_{n-1} + \Phi_n^{-1}\boldsymbol{x}_n y_n \\ &= \Phi_{n-1}^{-1}\boldsymbol{p}_{n-1} - \boldsymbol{k}_n\boldsymbol{x}_n^T\Phi_{n-1}^{-1}\boldsymbol{p}_{n-1} + \Phi_n^{-1}\boldsymbol{x}_n y_n \\ &= \boldsymbol{\theta}_{n-1} - \boldsymbol{k}_n\boldsymbol{x}_n^T\boldsymbol{\theta}_{n-1} + \Phi_n^{-1}\boldsymbol{x}_n y_n\end{aligned}$$

At this point we just need to prove that $\Phi_n^{-1}\boldsymbol{x}_n = \boldsymbol{k}_n$, since we would then be able to factorize and make e_n appear. This can be done by manipulating the expression for $\boldsymbol{k}_n$:

$$\begin{aligned}\boldsymbol{k}_n &= \frac{\beta^{-1}\Phi_{n-1}^{-1}\boldsymbol{x}_n}{1 + \beta^{-1}\boldsymbol{x}_n^T\Phi_{n-1}^{-1}\boldsymbol{x}_n} \\ \Rightarrow \boldsymbol{k}_n(1 + \beta^{-1}\boldsymbol{x}_n^T\Phi_{n-1}^{-1}\boldsymbol{x}_n) &= \beta^{-1}\Phi_{n-1}^{-1}\boldsymbol{x}_n \\ \Rightarrow \boldsymbol{k}_n &= \beta^{-1}\Phi_{n-1}^{-1}\boldsymbol{x}_n - \beta^{-1}\boldsymbol{k}_n\boldsymbol{x}_n^T\Phi_{n-1}^{-1}\boldsymbol{x}_n \\ &= (\beta^{-1}\Phi_{n-1}^{-1} - \beta^{-1}\boldsymbol{k}_n\boldsymbol{x}_n^T\Phi_{n-1}^{-1})\boldsymbol{x}_n \\ &= \Phi_n^{-1}\boldsymbol{x}_n\end{aligned}$$

where we have used once again the expression of Φ_n^{-1} found in the previous question.

We thus obtain the desired result:

$$\begin{aligned}\boldsymbol{\theta}_n &= \boldsymbol{\theta}_{n-1} - \boldsymbol{k}_n\boldsymbol{x}_n^T\boldsymbol{\theta}_{n-1} + \boldsymbol{k}_n y_n \\ &= \boldsymbol{\theta}_{n-1} + \boldsymbol{k}_n(y_n - \boldsymbol{x}_n^T\boldsymbol{\theta}_{n-1}) \\ &= \boldsymbol{\theta}_{n-1} + \boldsymbol{k}_n e_n\end{aligned}$$

which is the weight update for RLS.