Recursive Least Squares

İlker Bayram
ibayram@ieee.org

Introduction

Suppose X is a zero-mean Gaussian random vector with covariance Σ_0, and we make observations of the form

$$Z_k = h_k^T X + N_k, \tag{1}$$

where N_i's are independent zero-mean Gaussian random variables with variance σ^2, and h_k's are given constant observation vectors (generalization to matrices is straightforward). Our goal is to estimate X given $Z_1, \ldots, Z_k$, in a computationally efficient manner.

Thanks to the problem setup, all of the random variables are zero-mean Gaussian, and therefore, it follows that the unbiased minimum variance estimator of X in terms of $Z_1, \ldots, Z_k$, namely $\mathbb{E}(X|Z_1, Z_2, \ldots, Z_k)$, is a linear combination of Z_i's. We specifically want to update our estimate in an online fashion, as we make more and more observations, and use all the history, but without keeping all of the observations in the memory. The RLS algorithm provides a solution for this problem.

In the following, I will first give a characterization of the solution. Then, I will introduce an equivalent problem, which is usually used to motivate RLS, that does not refer to random variables. After that, I will derive the RLS algorithm, using the stochastic framework outlined above.

Characterization of the Solution

Suppose we stack Z_i's to form a matrix $\mathbf{Z}_k = \begin{bmatrix} Z_1 & \ldots & Z_k \end{bmatrix}^T$. Considering X to be a length-m random (column) vector, the problem we are trying to solve is

$$\min_{A \in \mathbb{R}^{m \times k}} \mathbb{E}\Big(\|X - A\ \mathbf{Z}_k\|_2^2 \Big). \tag{2}$$

After some algebra, we can write down the optimal A as

$$\hat{A} = \mathbb{E}(X\ \mathbf{Z}_k^T)\ \mathbb{E}(\mathbf{Z}_k\ \mathbf{Z}_k^T)^{-1}. \tag{3}$$

In order to write down an expression for these expected values, set $H_k = \begin{bmatrix} h_1 & h_2 & \dots & h_k \end{bmatrix}^T$. Then, we find

$$\mathbb{E}(X\ \mathbf{Z}_k^T) = \Sigma_0\, H_k^T \tag{4}$$

$$\mathbb{E}(\mathbf{Z}\ \mathbf{Z}^T) = H_k \Sigma_0\, H_k^T + \sigma^2\, I. \tag{5}$$

In principle then, the problem reduces to keeping track of the inverse of the matrix $H_k \Sigma_0\, H_k^T + \sigma^2\, I$ as k increases. In the following, we follow an equivalent but different course to derive RLS. Before that, though, I discuss an equivalent formulation that is more frequently used to when introducing RLS.

'Deterministic Setting'

An equivalent formulation can be obtained by making use of the properties of Gaussian random variables. For that, suppose the random experiment responsible for producing X along with noise terms N_k has been performed, and we observed $Z_i = z_i$, for $1 \leq i \leq k$. We would like to find the actual realization of X. Adopting the expected square error as the loss to minimize, the solution is $\mathbb{E}(X|Z_1 = z_1, \dots, Z_k = z_k)$. Thanks to all of the random variables being Gaussian, this expected value is actually the only mode of the posterior pdf $f_{X|Z_1,\dots Z_k}(x|z_1,\dots,z_k)$. Therefore, it suffices to look for the maximizer of this posterior pdf.

Note now that the posterior pdf satisfies

$$f_{X|Z_1,\dots Z_k}(x|z_1,\dots,z_k) \propto f_{Z_1,\dots Z_k|X}(z_1,\dots,z_k|x)\, f_X(x). \tag{6}$$

Taking logarithms, and plugging in the expressions for the pdfs, we find

$$\arg\max_x\ f_{X|Z_1,\dots Z_k}(x|z_1,\dots,z_k) \tag{7}$$

$$= \arg\min_x - \log\Big(f_{Z_1,\dots Z_k|X}(z_1,\dots,z_k|x) f_X(x) \Big) \tag{8}$$

$$= \arg\min_x\ \frac{1}{2\sigma^2} \sum_{i=1}^{k} (z_i - h_i^T\, x)^2 + \frac{1}{2}\, x^T \Sigma_0^{-1} x. \tag{9}$$

Solving for the minimizer, we find (using the definition of H_k above)

$$\hat{x} = \left(\frac{1}{\sigma^2} H^T\, H + \Sigma_0^{-1} \right)^{-1}\ H^T\ \begin{bmatrix} z_1 \\ \vdots \\ z_k \end{bmatrix}. \tag{10}$$

Note that the matrix $\hat{A} = (H^T\, H/\sigma^2 + \Sigma_0^{-1})^{-1}\, H^T$ estimating $\hat{x}$ from the observations appears to be different than $A = \Sigma_0\, H_k^T\, (H_k \Sigma_0\, H_k^T + \sigma^2\, I)^{-1}$ we derived above. However, they are in fact equal. To see this, multiply both matrices on the right by $C = (H^T\, H/\sigma^2 + \Sigma_0^{-1})$. This gives,

$$C\,\hat{A} = H^T \tag{11}$$

$$C\,A = \frac{1}{\sigma^2}\,H_k^T\,H_k\Sigma_0\,H_k^T\,(H_k\Sigma_0\,H_k^T + \sigma^2\,I)^{-1} + H_k^T\,(H_k\Sigma_0\,H_k^T + \sigma^2\,I)^{-1} \tag{12}$$

$$= \frac{1}{\sigma^2}\,H_k^T\,(H_k\Sigma_0\,H_k^T + \sigma^2\,I)^{-1}\,H_k\Sigma_0\,H_k^T + H_k^T\,(H_k\Sigma_0\,H_k^T + \sigma^2\,I)^{-1} \tag{13}$$

$$= H_k^T\,(H_k\Sigma_0\,H_k^T + \sigma^2\,I)^{-1}\,(H_k\Sigma_0\,H_k^T + \sigma^2\,I) \tag{14}$$

$$= H_k^T \tag{15}$$

Thus, both approaches yield the same solution, as expected.

The Algorithm

We finally consider the derivation of the RLS algorithm in this section.

Let

$$\hat{X}_k = \mathbb{E}(X|Z_1, Z_2, \ldots, Z_k) \tag{16}$$

denote the optimal estimator of X given $Z_1, \ldots, Z_k$, and define an error sequence

$$E_k = X - \hat{X}_k. \tag{17}$$

Let us denote the covariance of E_k for $1 \le i \le k$ by Σ_k. Also, let

$$C_k = \mathbb{E}(X^T\,E_k). \tag{18}$$

Suppose now that we receive the observation Z_{k+1}. RLS provides expressions for $\hat{X}_{k+1}$, Σ_{k+1}, and C_{k+1}, starting from $\hat{X}_k$, Σ_k, and C_k.

First, observe that

$$Z_{k+1} = h_{k+1}^T\,(\hat{X}_k + E_k) + N_{k+1}, \tag{19}$$

so that

$$Z_{k+1} - h_{k+1}^T\,\hat{X}_k = h_{k+1}^T\,E_k + N_{k+1}. \tag{20}$$

But $\mathbb{E}(E_k\,Z_i) = 0$ for $i \le k$, by the orthogonality principle. Also, since $\mathbb{E}(N_{k+1}\,Z_i) = 0$ for $i \le k$, the random variable $\tilde{Z}_{k+1}$ defined as

$$\tilde{Z}_{k+1} := Z_{k+1} - h_{k+1}^T\,\hat{X}_k \tag{21}$$

$$= h_{k+1}^T\,E_k + N_k \tag{22}$$

is independent of Z_i's. As a consequence,

$$
\begin{aligned}
\hat{X}_{k+1} &= \mathbb{E}(X|Z_{k+1}, Z_k, \ldots, Z_1) && (23)\\
&= \mathbb{E}(X|\tilde{Z}_{k+1}, Z_k, \ldots, Z_1) && (24)\\
&= \mathbb{E}(X|\tilde{Z}_{k+1}) + \mathbb{E}(X|Z_k, \ldots, Z_1) && (25)\\
&= \mathbb{E}(X|\tilde{Z}_{k+1}) + \hat{X}_k. && (26)
\end{aligned}
$$

Thanks to normality of the random variables, we can write

$$
\mathbb{E}(X|\tilde{Z}_{k+1}) = \mathbb{E}(X\, \tilde{Z}_{k+1})\ \mathbb{E}\big((\tilde{Z}_{k+1})^2\big)^{-1}\, \tilde{Z}_{k+1}. \tag{27}
$$

We need the two expected values, $\mathbb{E}(X\, \tilde{Z}_{k+1})$, and $\mathbb{E}\big((\tilde{Z}_{k+1})^2\big)$.

First note that

$$
\mathbb{E}(X\, \tilde{Z}_{k+1}) = \mathbb{E}(X\, E_k^T\, h_{k+1}) + \mathbb{E}(X\, N_{k+1}) = C_k\, h_{k+1}. \tag{28}
$$

We also have,

$$
\mathrm{var}(\tilde{Z}_{k+1}) = h_{k+1}^T\, \Sigma_1\, h_{k+1} + \sigma^2. \tag{29}
$$

So, the update equations are

$$
\hat{X}_{k+1} = C_k\, h_{k+1}(h_{k+1}^T\, \Sigma_k\, h_{k+1} + \sigma^2)^{-1}\, (Z_{k+1} - h_{k+1}^T\, \hat{X}_k) + \hat{X}_k \tag{30}
$$

In order to implement this, we also need update equations for C_k and Σ_k.

Now define $F_{k+1} = C_k\, h_{k+1}(h_{k+1}^T\, \Sigma_k\, h_{k+1} + \sigma^2)^{-1}$ so that,

$$
\begin{aligned}
\hat{X}_{k+1} &= F_{k+1}\, (Z_{k+1} - h_{k+1}^T\, \hat{X}_k) + \hat{X}_k && (31)\\
&= F_{k+1}\, (h_{k+1}^T\, E_k + N_{k+1}) + \hat{X}_k. && (32)
\end{aligned}
$$

where we used (20) in the last line. Subtracting this from X, we obtain a recursion equation for E_k as

$$
\begin{aligned}
E_{k+1} &= E_k - F_{k+1}\, (h_{k+1}^T\, E_k + N_{k+1}) && (33)\\
&= (I - F_{k+1}\, h_{k+1}^T)E_k - F_{k+1}\, N_{k+1}. && (34)
\end{aligned}
$$

It is now easier to see that

$$
C_{k+1} = \mathbb{E}(X\, E_{k+1}^T) = C_k(I - h_{k+1}\, F_{k+1}^T), \tag{35}
$$

and

$$
\Sigma_{k+1} = \mathbb{E}(E_{k+1}\, E_{k+1}^T) = (I - F_{k+1}\, h_{k+1}^T)\Sigma_k\, (I - h_{k+1}\, F_{k+1}^T) + \sigma^2\, F_{k+1}\, F_{k+1}^T. \tag{36}
$$