which depends \textit{linearly} on \(x_i\) (where const. does not depend on \(k\)). Although equivalent, the linear discriminant \(d_{\text{LINEAR}}\) can be far more efficient to compute than the quadratic \(d_{\text{MD}}\).

### 5.4 Combining Multiple Attack Traces

We have to combine the \(n_a\) individual leakage traces \(x_i\) from \(X_{k^*}\) into the final discriminant score \(d(k \mid X_{k^*})\). We present two sound options for doing so:

\textbf{Option 1:} Average all the traces in \(X_{k^*}\) (similar to the mean computation in (1)) in order to remove as much noise as possible and then use this single mean trace \(\bar{x}_{k^*}\) to compute

\[
    d_{\text{avg}}(k \mid X_{k^*}) = d(k \mid \bar{x}_{k^*}). \tag{26}
\]

This option is computationally fast, requiring \(O(n_a + m^3)\) time for any presented discriminant, but it does not use all the information from the available attack traces (in particular the noise).

\textbf{Option 2:} Compute the joint likelihood \(l(k \mid X_{k^*}) = \prod_{x_i \in X_{k^*}} l(k \mid x_i)\). By applying the logarithm to both sides we have \(\log l(k \mid X_{k^*}) = \sum_{x_i \in X_{k^*}} \log l(k \mid x_i)\) and we obtain the derived scores:

\[
    d_{\text{LOG}}^{\text{joint}}(k \mid X_{k^*}) = -\frac{n_a}{2} \log |S_k| - \frac{1}{2} \sum_{x_i \in X_{k^*}} (x_i - \bar{x}_k)'S_k^{-1}(x_i - \bar{x}_k), \tag{27}
\]

\[
    d_{\text{MD}}^{\text{joint}}(k \mid X_{k^*}) = -\frac{1}{2} \sum_{x_i \in X_{k^*}} (x_i - \bar{x}_k)'S_k^{-1}(x_i - \bar{x}_k), \tag{28}
\]

\[
    d_{\text{LINEAR}}^{\text{joint}}(k \mid X_{k^*}) = \bar{x}_k'S_{\text{pooled}}^{-1}\left(\sum_{x_i \in X_{k^*}} x_i\right) - \frac{n_a}{2} \bar{x}_k'S_{\text{pooled}}^{-1}\bar{x}_k. \tag{29}
\]

Given the \(n_a\) leakage traces \(x_i \in X_{k^*}\), \(d_{\text{LOG}}\) and \(d_{\text{MD}}\) require time \(O(n_a m^3)\) while \(d_{\text{LINEAR}}\) only requires \(O(n_a + m^3)\), since the operations \(\bar{x}_k'S_{\text{pooled}}^{-1}\) and \(\bar{x}_k'S_{\text{pooled}}^{-1}\bar{x}_k\) only need to be done once, which is a great advantage in practice. As a practical example, our evaluations of the guessing entropy (see Sect. 6) for \(m = 125\) and \(1 \leq n_a \leq 1000\) took about 3.5 days with \(d_{\text{LOG}}\) but only 30 min with \(d_{\text{LINEAR}}.\)

\(^9\) We note that for \(d_{\text{LINEAR}}\) the computation time is the same regardless of which option we use to combine the traces, and both give the same results for the template attack.

\(^9\) MATLAB, single core CPU with 3794 MIPS.