# The proof of Least Square Learning Rule

Let $\varepsilon_k = T_{K-}Y_k$ ,we use Least Mean Square Error to ensure the minimum total error. As long as the total error approaches zero, the best solution is found. Therefore, we are looking for the minimum of $\langle \varepsilon_k^2 \rangle$。

Proof:

mean→ $\langle \varepsilon_k^2 \rangle = \frac{1}{L}\sum_{k=1}^{L}\varepsilon_k^2 = \frac{1}{L}\sum_{k=1}^{L}(T_k - Y_k)^2 = \frac{1}{L}\sum_{k=1}^{L}(T_k^2 - 2T_kY_k + Y_k^2)$

$= \frac{1}{L}\sum_{k=1}^{L}T^2 - \frac{2}{L}\sum_{k=1}^{L}T_kY_k + \frac{1}{L}\sum_{k=1}^{L}Y_k^2$ → let $\langle T_k^2 \rangle$ is mean of

$= \langle T_k^2 \rangle - \frac{2}{L}\sum_{k=1}^{L}T_kY_k + \frac{1}{L}\cdot[W^t(\sum_{k=1}^{L}X_kX_k^t)W]$

$$\boxed{\begin{aligned}\sum_{k=1}^{1}Y_k^2 &= \sum_{k=1}^{L}(\sum_{i=1}^{n}W_iX_{ik})^2 = \sum_{k=1}^{L}(W^tX_k)^2 = \sum_{k=1}^{L}(W^tX_k)(X_k^tW)\\ &= W^t(\sum_{k=1}^{L}X_k\cdot X_k^t)\cdot W\end{aligned}}$$

$= \langle T_k^2 \rangle - \frac{2}{L}\sum_{k=1}^{L}T_kY_k + W^t[\frac{1}{L}(\sum_{k=1}^{L}X_kX^t)]W$

$= \langle T_k^2 \rangle - \frac{2}{L}\sum_{k=1}^{L}T_k(W_k^tX_k) + W^t\langle X_kX_k^t\rangle W$

$= \langle T_k^2 \rangle - 2[\frac{1}{L}\sum_{k=1}^{L}(T_kX_k^t)W] + W^t\langle X_kX_k^t\rangle W$

$= \langle T_k^2 \rangle - 2\langle T_kX_k^t\rangle W + W^t\langle X_kX_k^t\rangle W$

Let $R_k = X_kX_k^t$, i.e., $R_k$ is a n×n matrix，also called Correlation Matrix

Let $R'$=R1+R2+…+RK+…+RL$=\sum_{k=1}^{L}T_kX_k^t$

Let $R = \frac{R'}{L}$ （i.e. mean of $R'$ ）⇒which is R=$\langle X_kX_k^t\rangle$

$= \langle T_k^2 \rangle + W^tRW - 2\langle T_kX_k^t\rangle$ W

★Find W such that $\langle \varepsilon_k^2 \rangle$ is minimal

$\frac{\partial\langle\varepsilon_k^2\rangle}{\partial W} = [\langle T_k^2\rangle + W^tRW - 2\langle T_kX_k^t\rangle W]'$

=2RW－2$\langle T_kX_k\rangle$ =2RW-2P Let P=$\langle T_kX_k^t\rangle$

if $\frac{\partial\langle\varepsilon_k^2\rangle}{\partial W} = \phi \Rightarrow$ 2RW*－2P=$\phi$ 即 W*=R-1P 或 RW*=P