Predicción Lineal

Preparado por
Dr. Matías Zañartu, e-mail: Matias.Zanartu@usm.cl

I. Introducción

La codificación predictiva lineal (LPC), es un tipo de codificación utilizado ampliamente en aplicaciones de audio digital. Por ejemplo, en sistemas de procesado de voz, se usa partiendo de la idea de que la voz puede modelarse como una combinación lineal de P muestras anteriores más una señal de error.

II. Predicción Lineal

Supongamos que la señal discreta que utilizamos (ejemplo, voz) corresponde a un proceso estocástico discreto, donde cada muestra es una variable aleatoria y la secuencia $\{..., x_{-1}, x_0, x_1, x_2, ...\}$ representa a la señal.

Observación.

- Se usa x_i y no $x(i)$ por notación para evidenciar que cada muestra es una variable independiente.
- LPC busca encontrar alguna correlación entre las muestras (lineal y de bajo orden).

Luego, la aproximación lineal predictiva se puede representar como $\hat{x}_n$, donde

$$\hat{x}_n = \sum_{k=1}^{P} a_k x_{n-k}, \tag{1}$$

$$= \mathbf{a}^T \mathbf{x_n}, \tag{2}$$

donde (2) representa el producto punto entre dos vectores, resultando un escalar. La notación en negrita se asocia a vectores o matrices. Los vectores $\mathbf{a}$ y $\mathbf{x_n}$, son vectores columna de largo P definidos como

$$\mathbf{a} = [a_1 \ a_2 \ ... \ a_P]^T, \tag{3}$$

$$\mathbf{x_n} = [x_{n-1} \ x_{n-2} \ ... \ x_{n-P}]^T. \tag{4}$$

Luego, el error de predicción está dado por

$$e_n = x_n - \hat{x}_n, \tag{5}$$

$$= x_n - \mathbf{a}^T \mathbf{x_n}. \tag{6}$$

Pregunta.
¿Cómo obtener coeficientes óptimos?. Una alternativa es usar MMSE (Minimum mean square error) que minimiza $E[e_n^2]$, es decir buscar el vector $\mathbf{a}$ que hacen que $\nabla E[e_n^2] = 0$

Observación.
Recuerden que $E[x] = \sum_{i=-\infty}^{\infty} x_i p_i$, donde x_i corresponde a los valores o estados de x en el tiempo i y p_i es la probabilidad asociada a la secuencia de dicho valor. Si la probabilidad de cada estado es igual a $p_i = \frac{1}{N}$, lo que hace que $E[x]$ sea equivalente a tomar un promedio aritmético.

Luego, calculando el MMSE se tiene,

$$
\begin{aligned}
E[e_n^2] &= E[(x_n - \hat{x}_n)^2], && (7)\\
&= E[(x_n - \mathbf{a}^T\mathbf{x_n})^2], && (8)\\
&= E[x_n^2 - 2x_n\mathbf{a}^T\mathbf{x_n} + (\mathbf{a}^T\mathbf{x_n})(\mathbf{a}^T\mathbf{x_n})], && (9)\\
&= E[x_n^2] - 2\mathbf{a}^T E[x_n\mathbf{x_n}] + \mathbf{a}^T E[\mathbf{x_n}\mathbf{x_n}^T]\mathbf{a}, && (10)
\end{aligned}
$$

donde

$$
E[x_n\mathbf{x_n}] = \begin{bmatrix} E[x_n x_{n-1}] \\ E[x_n x_{n-2}] \\ \vdots \\ E[x_n x_{n-P}] \end{bmatrix} = \begin{bmatrix} r_{xx}(1) \\ r_{xx}(2) \\ \vdots \\ r_{xx}(P) \end{bmatrix} = \mathbf{r_x}, \qquad (11)
$$

y

$$
\begin{aligned}
E[\mathbf{x_n}\mathbf{x_n}^T] &= E\begin{bmatrix} x_{n-1}x_{n-1} & x_{n-1}x_{n-2} & \dots & x_{n-1}x_{n-P} \\ x_{n-2}x_{n-1} & x_{n-2}x_{n-2} & \dots & x_{n-2}x_{n-P} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-P}x_{n-1} & x_{n-P}x_{n-2} & \dots & x_{n-P}x_{n-P} \end{bmatrix} && (12)\\
&= \begin{bmatrix} r_{xx}(0) & r_{xx}(1) & \dots & r_{xx}(P-1) \\ r_{xx}(1) & r_{xx}(0) & \dots & r_{xx}(P-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_{xx}(P-1) & r_{xx}(P-2) & \dots & r_{xx}(0) \end{bmatrix} = \mathbf{R_x}, && (13)
\end{aligned}
$$

donde $\mathbf{R_x}$ es simétrica y Toeplitz.

Observación.

- $E[x_n x_m] = r_{xx}(n,m)$ (autocorrelación)
- Si x es estacionario (WSS), entonces $r_{xx}(n,m) = r_{xx}(\mid n-m \mid)$
- Se asume en este caso que la voz es WSS en intervalos pequeños de tiempo

Luego,

$$
\begin{aligned}
E[e_n^2] &= E[x_n^2] - 2\mathbf{a}^T E[x_n\mathbf{x_n}] + \mathbf{a}^T E[\mathbf{x_n}\mathbf{x_n}^T]\mathbf{a}, && (14)\\
&= E[x_n^2] - 2\mathbf{a}^T\mathbf{r_x} + \mathbf{a}^T\mathbf{R_x}\mathbf{a}, && (15)
\end{aligned}
$$

luego se debe buscar aquellos coeficientes $\hat{\mathbf{a}}$ que hacen $\nabla E[e_n^2] = 0$, es decir

$$
\nabla_{\mathbf{a}} E[e_n^2] = -2\mathbf{r_x} + 2\mathbf{R_x}\hat{\mathbf{a}} = \mathbf{0}, \qquad (16)
$$

entonces

$$
\mathbf{R_x}\hat{\mathbf{a}} = \mathbf{r_x}, \qquad (17)
$$

dada la estructura de $\mathbf{R_x}$, la ecuación (17) se resuelve eficientemente usando algoritmos tales como Levison-Durbin, el cual es muy útil en tiempo real.

Pregunta.
¿Cómo estimar $\mathbf{r_x}$?. Una forma es considerar un segmento de la señal (donde puede ser considerada WSS) de largo N y usar

$$
\hat{\mathbf{r}}_{xx}(m) = \frac{1}{N}\sum_{n=0}^{N-m-1} x_n x_{n+m}, \quad 0 \le m \le P \qquad (18)
$$

Observación.

- MatLab permite obtener cada $\hat{\mathbf{r}}_{xx}(m)$ (y por ende $\mathbf{r_x}$) usando el comando $xcorr$
- MatLab genera usando $\mathbf{R_x}$ mediante $toeplitz$

Finalmente, de las muestras de x, determinar $\mathbf{r_x}$ y $\mathbf{R_x}$, luego resolver la ecuación (17) y encontrar a. Con los coeficientes, es posible modelar el efecto del tracto vocal y sintetizar sonidos de voz.