TRATAMIENTO DIGITAL DE SEÑALES
Ingeniería de Telecomunicación (4º, 2º c)

Unidad 6ª: Estimación y regresión lineales

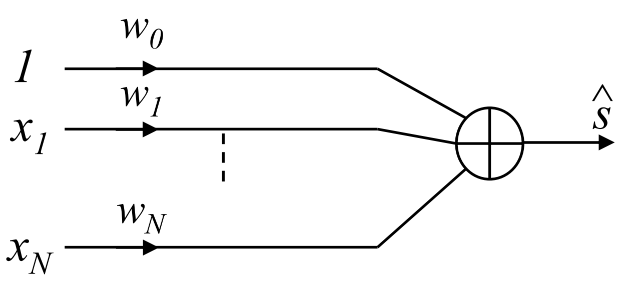
Aníbal R. Figueiras Vidal
Jesús Cid Sueiro
Ángel Navia Vázquez

Área de Teoría de la Señal y Comunicaciones
Universidad Carlos III de Madrid

ación Lineal Cuadrático Medio Mínima)

Como en el caso de los decisores, los estimadores lineales proporcionan s por su sencillez y facilidad de interpretación, así que es frecuente r a ellos. (Además, son estrictamente óptimos en algunas situaciones, ara el Problema General Gaussiano).

Supóngase que la va s se estima a partir de las N vas $\boldsymbol{x}$ *imponiendo la lineal* $\hat{s} = w_0 + \boldsymbol{w}^T \boldsymbol{x} = \boldsymbol{w}_e^T \boldsymbol{x}_e \quad \left(\boldsymbol{x}_e = \left[1 \ \boldsymbol{x}^T \right]^T \right)$



ínese $\quad \hat{\boldsymbol{w}}_e : \min_{\boldsymbol{w}_e} E\left\{ \left(s - \hat{s} \right)^2 \right\} = \min_{\boldsymbol{w}_e} E\left\{ e^2 \right\}$

$$e^2 = \left(s - \mathbf{w}_e^T \mathbf{x}_e\right)^2; \quad E\left\{e^2\right\} = E\left\{\left(s - \mathbf{w}_e^T \mathbf{x}_e\right)^2\right\}$$

$$\frac{\partial E\left\{e^2\right\}}{\partial w_n} = E\left\{\frac{\partial e^2}{\partial w_n}\right\} = E\left\{\frac{\partial e^2}{\partial e}\frac{\partial e}{\partial w_n}\right\} = 2E\left\{e\frac{\partial\left(s - \mathbf{w}_e^T \mathbf{x}_e\right)}{\partial w_n}\right\}$$

ualaremos a 0

a w_0, resulta: $E\{e\}\big|_{\mathbf{w}_e = \hat{\mathbf{w}}_e} = 0$

$$E\{s\} - \hat{w}_0 - \hat{\mathbf{w}}^T E\{\mathbf{x}\} = 0$$

$$\hat{w}_0 = E\{s\} - \hat{\mathbf{w}}^T E\{\mathbf{x}\}$$

a los demás pesos: $E\{e x_n\}\big|_{w = \hat{w}_e} = 0$

en forma bloque: $E\left\{\hat{e}\,\mathbf{x}\right\} = \mathbf{0} = E\left\{\left(s - \hat{\mathbf{w}}_e^T \mathbf{x}_e\right)\mathbf{x}\right\}$

Principio de Ortogonalidad: $\hat{e} \perp x_i, \quad i = 1, \ldots, N$

(también en versión extendida)

nprueba fácilmente que es un mínimo, al ser la matriz de derivadas as un producto externo).

cipio de Ortogonalidad para la formulación extendida conduce a:

$$
\begin{array}{lllllll}
+\hat{w}_1 E\{x_1\} & +\hat{w}_2 E\{x_2\}+ & \cdots & +\hat{w}_N E\{x_N\} & = & E\{s\} \\
+\hat{w}_1 E\{x_1x_1\} & +\hat{w}_2 E\{x_2x_1\}+ & \cdots & +\hat{w}_N E\{x_Nx_1\} & = & E\{sx_1\} \\
+\hat{w}_1 E\{x_1x_2\} & +\hat{w}_2 E\{x_2x_2\}+ & \cdots & +\hat{w}_N E\{x_Nx_2\} & = & E\{sx_2\} \\
 & & & & & \vdots \\
+\hat{w}_1 E\{x_1x_N\} & +\hat{w}_2 E\{x_2x_N\}+ & \cdots & +\hat{w}_N E\{x_Nx_N\} & = & E\{sx_N\}
\end{array}
$$

ede reescribirse utilizando las covarianzas v : $E\{x_ix_j\} = r_{x_ix_j} =$

$+ E\{x_i\}E\{x_j\}$, *y análogamente para* $E\{sx_i\}$;

e entonces en la ecuación j+1

ando de la forma $\left(\hat{w}_0 + \hat{w}_1 E\{x_1\} + \hat{w}_2 E\{x_2\} + \ldots + \hat{w}_N E\{x_N\}\right) E\{x_j\}$

rmino de la izquierda;

ando de la forma $E\{s\}E\{x_j\}$ *en el de la derecha;*

l ser iguales (por la primera ecuación), pueden eliminarse; y debe erse así, puesto que se eliminan términos de ***colinealidad*** *en las ones, y con ello se alivian los problemas numéricos en la solución.*

Así, se resolverá por un lado la primera ecuación, y, por otro, las as ***Ecuaciones Normales*** *(N ecuaciones con N incógnitas)*

$$
\begin{array}{ccccccccc}
\hat{w}_1 v_{x_1x_1} & + & \hat{w}_2 v_{x_2x_1} & + & \cdots & + \hat{w}_N v_{x_Nx_1} & = & v_{sx_1} \\
\hat{w}_1 v_{x_1x_2} & + & \hat{w}_2 v_{x_2x_2} & + & \cdots & + \hat{w}_N v_{x_Nx_2} & = & v_{sx_2} \\
\vdots & & & & & & & \vdots \\
\hat{w}_1 v_{x_1x_N} & + & \hat{w}_2 v_{x_2x_N} & + & \cdots & + \hat{w}_N v_{x_Nx_N} & = & v_{sx_N}
\end{array}
$$

n las únicas a resolver si ***x*** *y s tienen medias nulas).*

En forma matricial

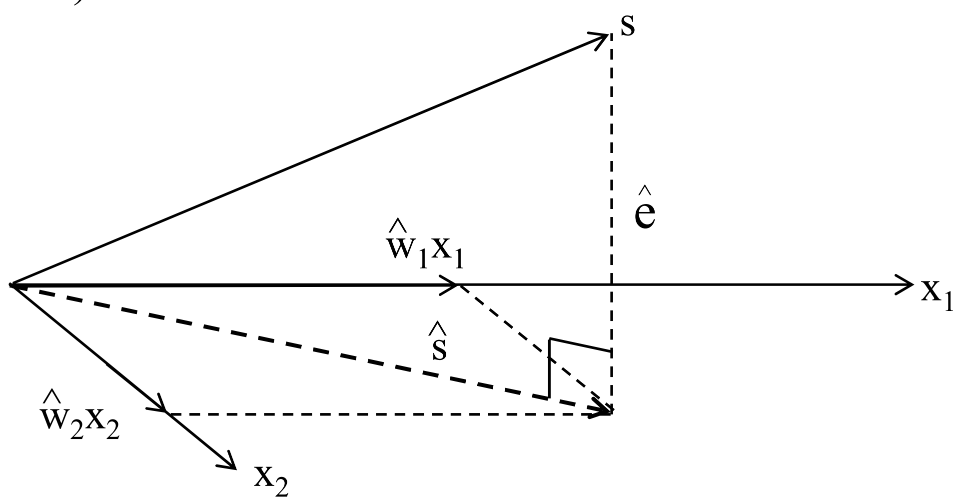
$$\boldsymbol{V}_{xx}\hat{\boldsymbol{w}} = \boldsymbol{v}_{sx} \qquad \textit{(nótese: } [\boldsymbol{V}_{xx}]_{i,j} = v_{x_jx_i} \textit{)}$$

lución es

$$\hat{\boldsymbol{w}} = \boldsymbol{V}_{xx}^{-1}\boldsymbol{v}_{sx}$$

<u>...etación</u>

...cipio de Ortogonalidad tiene una interpretación obvia considerando que las ...ueden representar como vectores con módulo igual a la raíz cuadrada de su ...a (medias nulas):



...ño del error es mínimo si s se proyecta ortogonalmente sobre el subespacio ...do por las **x**; y el estimador lineal buscado no es otra cosa que la **...ción resultante**, ŝ. Por ello, se habla también de **Teorema de Proyección**.

...Debe notarse que $E\{\hat{e}\hat{s}\} = E\{\hat{e}\hat{\mathbf{w}}^T\mathbf{x}\} = 0$

ios de discusión

es la contribución a la reducción del error de cada variable x_i? ndase de las medias).

n la misma figura que se ha utilizado para ilustrar el Principio de nalidad se observa que la intervención de x_1 en el valor de $\hat{e}$ (y de $\hat{s}$)... e de cómo sea x_2.

nalíticamente:

$$\hat{e}^2\} = E\{(s - \hat{\boldsymbol{w}}^T \boldsymbol{x})\hat{e}\} = E\{s\,\hat{e}\} = E\{s(s - \hat{\boldsymbol{w}}^T \boldsymbol{x})\} = E\{s^2\} - \hat{\boldsymbol{w}}^T \boldsymbol{v}_{sx}$$

$$\{\hat{e}^2\} = E\{s^2\} - \boldsymbol{v}_{sx}^T \boldsymbol{V}_{xx}^{-1} \boldsymbol{v}_{sx}$$

queda de manifiesto lo dicho: y se percibe que la covarianza entre s y x_i ca directamente la importancia de x_i para reducir el error (salvo que se elegir una variable de entre dos).

Por tanto, si se tratase de seleccionar variables, no se puede hacer de o con sus covarianzas individuales con respecto a s. Tampoco de o con sus coeficientes $\hat{\mathbf{w}}$: ya que sus valores dependen de todas las variables.

*Naturalmente, en el caso de que las **x** sean **incorrelacionadas** (varianzas as nulas), $\mathbf{V}_{xx}$ se diagonaliza, y resulta manifiesto que cada variable de acuerdo con $v_{x_i x_i}^{-1} v_{sx_i}^{2}\left(w_i v_{sx_i}\right)$.*

*Y también es obvio que, si se trata de añadir una nueva variable a un dor lineal **fijado**, ha de elegirse la más correlacionada con el **error**.*

El problema de selección de variables es importante, como se sabe: debe arse que la utilización de variables linealmente dependientes convierte singular.

Métodos de selección de variables para estimación lms

...riables dato se pueden "ortogonalizar" (medias nulas) de acuerdo con el ...miento de Gram-Schmidt.

...ule algorítmicamente el procedimiento.

...l procedimiento útil para selección de variables?

...puesto que se ha fijado un orden $x_1, \ldots, x_N$, se trata de ir creando variables que ...n ortogonales a las anteriores: para ello se restan las partes no ortogonales

$$= x_1$$

$$= x_2 - \frac{E\{x_1' \, x_2\}}{E\{x_1' \, x_1'\}} x_1'$$

$$= x_3 - \frac{E\{x_1' \, x_3\}}{E\{x_1' \, x_1'\}} x_1' - \frac{E\{x_2' \, x_2\}}{E\{x_2' \, x_2'\}} x_2'$$

es decir :

$$x_1' = x_1$$

$$a_{ln} = \frac{E\{x_l' \, x_n\}}{E\{x_l' \, x_l'\}}, \qquad 1 \leq l \leq n$$

$$x_n' = x_n - \sum_{l=1}^{n-1} a_{ln} x_l', \qquad 2 \leq n \leq N$$

versiones algorítmicas de mejores características numéricas)

tá claro que es inmediata la selección de las x': pero ello no lleva plícita una selección de las x, ya que en las x' seleccionadas pueden trar cualesquiera combinaciones de las variables originales.

Sin embargo, sí es cierto que la aplicación de este proceso evita sufrir efectos de colinealidades entre variables).

Vo es solución ordenar la ortogonalización según la correlación de los tos con s, ya que no se puede prever el efecto que tiene el proceso sobre has variables (ni siquiera es óptimo elegir en cada paso la variable que lugar a la variable transformada de mayor efecto).

Pero también es verdad que el efecto favorable de evitar colinealidades rmite obtener buenos resultados aplicando estos procedimientos bóptimos: que se conocen como métodos de ***Ortogonalización de nimos Cuadrados*** *(OLS, "Orthogonal Least Squares") (el nombre se fiere, estrictamente, a su aplicación en forma* ***muestral****).*

variaría la formulación expuesta si se tratase con vas complejas?

Considerando que hay que manejar $E\{|e|^2\} = E\{ee^*\}$, se llega de ato a ver que hay que sustituir $E\{x_i x_j\}$ y $E\{sx_i\}$ por $E\{x_i x_j^*\}$ y $E\{sx_i^*\}$.

Conviene notar que la transposición se convierte en transposición ca: y que esta operación conjuga V_{xx}.

hay que extender la formulación si se estimase una variable mensional, **s**?

Bastaría proceder fila a fila para encontrar como solución

$$\hat{\mathbf{s}} = \hat{\mathbf{W}}^T \mathbf{x}$$

$$\hat{\mathbf{W}} = \mathbf{V}_{xx}^{-1} \mathbf{V}_{sx}$$

que $\mathbf{V}_{sx} = E\{\mathbf{x}\mathbf{s}^T\}$*)*

ıes analíticas y máquina

i la física del problema es conocida, V_{xx} y v_{sx} se pueden obtener de onocimiento: se trataría entonces de un aplicación **analítica**.

i no es así, tendrían que estimarse V_{xx} y v_{sx} a partir de observaciones **rales** $\{s^{(k)}, \mathbf{x}^{(k)}\}_{k=1}^{K}$: en cuyo caso se hablaría de una formulación **ıa**.

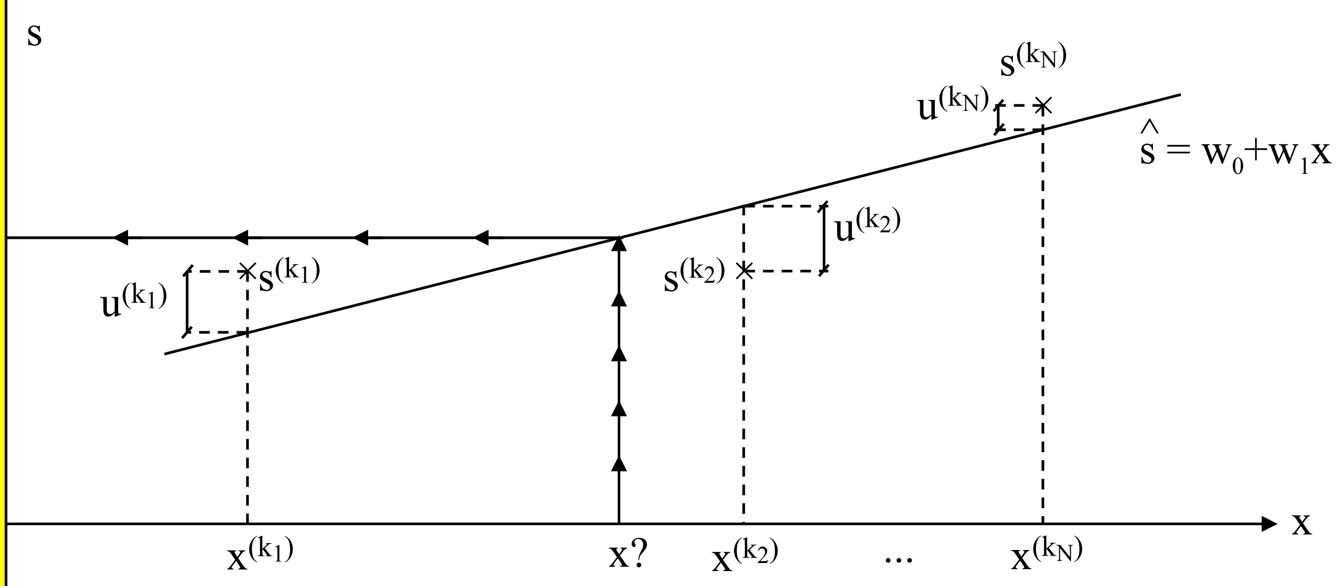
Desde ahora se hace notar que, para comprobar la coherencia de lo que se poniendo (el modelo físico, lo razonable de emplear la estimación lms), ıe, cuando sea posible, verificar que los resultados son estables para aproximaciones: analítica y máquina si hay modelo físico y datos bles, o entre varias aproximaciones máquina.

gresión Lineal

Este problema se formula análogamente a la estimación de la media de gaussiana: se admite que cada observación de s es una combinación le las variables deterministas $\{x_n\}$ y una componente aleatoria, llamada **o** o **innovación**, u, que se supone gaussiana de media cero: además, con ma varianza para cualesquiera valores de las $\{x_n\}$ (condición de **edasticidad**).

La va. se estima **por su media** (lo que minimiza el error cuadrático de la estimación); y ésta se obtiene a partir de un conjunto de aciones $\{s^{(k)}, \mathbf{x}^{(k)}\}_{k=1}^{K}$ (típicamente, K>>N), ajustando los parámetros vía **mínimos cuadrados** (LS).

...do que la Regresión Lineal pretende lo que se ilustra en la figura: para ...alor de x (**x**), establecer qué parte de s se debe al mismo, minimizando el ...de u.



ːl ajuste LS de los parámetros minimiza el error cuadrático total ıado: suma de los cuadrados de

$$u^{(k)} = s^{(k)} - \mathbf{w}_e^T \mathbf{x}_e^{(k)} \ , \qquad k = 1,\dots,K$$

ː, el módulo del vector

$$\bar{u} = \bar{s} - \mathbf{X}_e \mathbf{w}_e$$

$$\mathbf{X}_e = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \cdots & x_N^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \cdots & x_N^{(2)} \\ \vdots & & & & \\ 1 & x_1^{(K)} & x_2^{(K)} & \cdots & x_N^{(K)} \end{bmatrix} \qquad (K \times (N+1))$$

ıno es otra cosa que aproximar $\bar{s}$ mediante una combinación lineal de los vectores columna de $\mathbf{X}_e$, minimizando el error cuadrático: se aplicará el ıio de Ortogonalidad

$$\hat{\bar{u}} \perp \mathrm{col}\{\mathbf{X}_e\}$$

r, en forma bloque

$$\mathbf{X}_e^{T}\,\hat{\bar{u}} = \mathbf{0}$$

$$\mathbf{X}_e^{T}\bar{s} - \mathbf{X}_e^{T}\mathbf{X}_e\,\hat{\mathbf{w}}_e = \mathbf{0}$$

de, supuesta $\mathbf{X}_e^{T}\mathbf{X}_e$ invertible,

$$\hat{\mathbf{w}}_e = \left(\mathbf{X}_e^{T}\mathbf{X}_e\right)^{-1}\mathbf{X}_e^{T}\,\bar{s}$$

La matriz factor de $\bar{s}$ se denomina **seudoinversa de Moore-Penrose** , $\mathbf{X}_e^{\#}$ y es la solución del problema LS que aquí ha surgido. Proyecta $\bar{s}$ l subespacio formado por las columnas de $\mathbf{X}_e$.

ión

ómo procedería en una situación de heterocedasticidad, en que u tiene rianza dependiente de ***x****?*

La solución pasaría por convertir el problema en homocedástico: tribuyendo u.

Si la forma de la heterocedasticidad es conocida, se puede aplicar en la ación, y se llega a una solución LS ponderada; si no es así, han de rse aproximaciones sucesivas asumiendo homocedasticidad local, y ndo las varianzas a partir de los residuos.

ión Lineal vs Estimación Lineal

otar que las expresiones obtenidas para las soluciones son análogas:

$\mathbf{X}_e$ hace el papel de $\mathbf{V}_{xx}$

$\bar{s}$ hace el papel de v_{sx}

echo, es inmediato comprobar que la solución de la Estimación Lineal se rmalmente idéntica a la de la Regresión si se sigue la vía muestral y se n correlaciones de forma muestral para las ecuaciones extendidas.

Pero el planteamiento es conceptualmente distinto:

En E. L., las **x** son **va**

En R.L., las **x** son **deterministas**

o, por tanto, diferentes situaciones de aplicación.

Si cabe la opción de modelar las **x** como deterministas o como aleatorias, aro que la segunda opción será tanto más ventajosa cuanto **más miento** estadístico se tenga de las **x**; la primera, cuanto menos: por o, cuando los valores de las **x** puedan ser "controladas" externamente.

ɔs semilineales

Reciben este nombre todos los modelos construidos

ransformando no linealmente las variables **x** en otras variables **y**

ıtilizando para la estimación/regresión una combinación lineal de las **y**;

que, fijada la transformación, podemos proceder con las **y** de los modos
y se tendrán métodos **no lineales** de diseño sencillo, cuya eficacia
erá de lo adecuado de la **transformación** para el problema bajo estudio.

;on muy típicos los métodos semilineales basados en un previo **ımiento** de las muestras $\mathbf{x}^{(k)}$, empleando después como $y^{(k)}$ las ıdes de cada $\mathbf{x}^{(k)}$ con los representantes de los grupos, $\mathbf{m}_j$. Estudiaremos ıicas de agrupamiento en la Unidad 10.

'ambién se incluyen aquí los modelos **polinómicos**.

$$\hat{s} = w_o + \sum_{n=1}^{N} w_n x_n + \sum_{n=1}^{N} \sum_{n'=1}^{N} w_{nn'} x_n x_{n'} + \ldots$$

ıue los términos del polinomio son transformaciones de las $\{x_n\}$. Como , la dificultad aquí radica en que hay una **explosión dimensional** con el lel polinomio.

ıa técnica (subóptima) para construir modelos polinómicos es el llamado) de Manejo de Datos en Grupo ("Group Method Data Handling", H), debido a Ivakhnenko. Se basa en que multiplicando polinomios se sus grados, y que los polinomios de grado 2 en dos variables tienen sólo cientes; así,

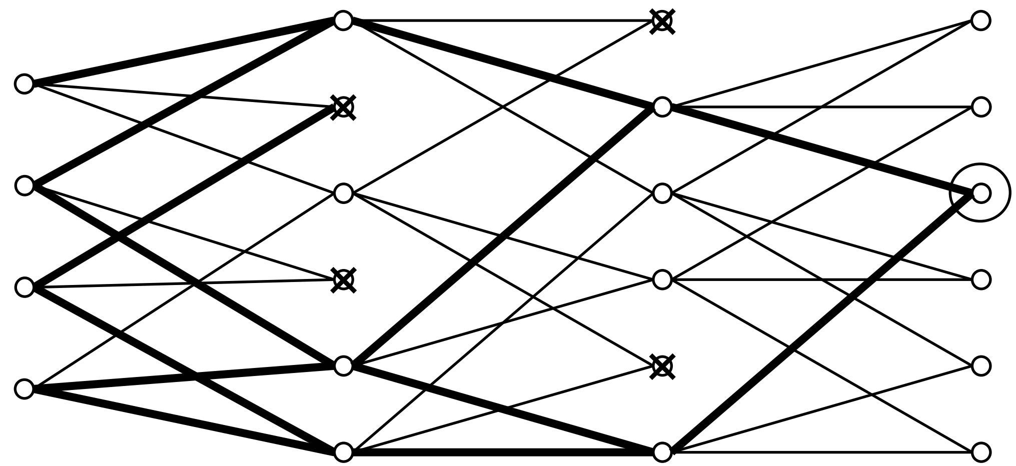
emplean todos los posibles términos de dos variables de orden 2 a partir las variables (hay $\binom{N+1}{2}$ términos así);

seleccionan los N "mejores";

itera el proceso, construyendo polinomios de dos variables y orden 2 a rtir de las salidas del paso anterior (el orden sube a 4, 6, 8, ...), hasta que saturan los resultados;

elige como solución la correspondiente a la mejor salida del último so, y se procede a un diseño final.

de ilustrar el proceso como sigue:



El GMDH tiene mecanismos que aseguran una buena generalización a iterativamente con una parte de las muestras y verifica sobre el resto), na gran cantidad de variantes.