

B3 - Apuntes de filtrado lineal

Nombre del curso: Teoría Moderna de la Detección y Estimación

Autores: Miguel Lázaro Gredilla



Contents

1	Estimación analítica	1
1.1	Visión general de los problemas de estimación	1
1.1.1	Estimadores de parámetro determinista y de variable aleatoria	2
1.1.2	Estimación analítica y máquina	4
1.2	Diseño analítico de estimadores de variable aleatoria. Teoría bayesiana de la estimación	4
1.2.1	Modelado estadístico de los problemas de estimación	4
1.2.2	Funciones de coste	5
1.2.3	Coste medio.	7
1.2.4	Estimador bayesiano.	8
1.3	Estimadores bayesianos de uso frecuente	9
1.3.1	Estimador de mínimo error cuadrático medio (MMSE)	9
1.3.2	Estimador de mínimo error absoluto (MAD)	11
1.3.3	Estimador de máximo a posteriori (MAP)	12
1.4	Estimación de máxima verosimilitud	14
1.5	Estimación con distribuciones gaussianas	17
1.5.1	Caso unidimensional	18
1.5.2	Caso con variables multidimensionales	20
1.6	Estimación con restricciones	21
1.6.1	Principios generales	21
1.6.2	Estimación lineal de mínimo error cuadrático medio	22
1.7	Caracterización de estimadores	28
1.7.1	Sesgo y varianza de estimadores de parámetros deterministas	28
1.7.2	Sesgo y varianza de estimadores de variables aleatorias	31
1.8	Apéndices	32
1.8.1	Casos particulares gaussianos	32
1.8.2	Principio de Ortogonalidad. Interpretación geométrica	34
1.9	Problemas	35
2	Aprendizaje Máquina	37
2.1	Principios generales del aprendizaje máquina	37
2.2	Métodos Paramétricos y no Paramétricos	38
2.3	Estimación Máquina No Paramétrica: Método del vecino más próximo ..	38
2.4	Estimación Máquina Paramétrica: Regresión de Mínimos Cuadrados	39
2.4.1	Modelos Semilineales	39

2.5	Generalización	39
3	Decisión analítica	41
3.1	Introducción al problema de decisión	41
3.1.1	Regiones de decisión	42
3.1.2	Diseño de decisores	43
3.2	Diseño analítico de decisores	44
3.2.1	Modelado estadístico de los problemas de decisión	44
3.2.2	Riesgo	44
3.2.3	Teoría bayesiana de la decisión	47
3.2.4	Decisión ML	50
3.3	Decisores binarios	51
3.3.1	Riesgo de un decisor binario	51
3.3.2	Función discriminante	52
3.3.3	Decisores binarios de mínimo riesgo	53
3.3.4	Decisor ML	54
3.3.5	Decisores no Bayesianos	57
3.4	El caso Gaussiano	62
3.4.1	Varianzas iguales	64
3.4.2	Medias nulas	65
3.5	Apéndices	67
3.5.1	Diseño analítico de decisores con costes dependientes de la observación	67
3.6	Problemas	69
4	Decisión máquina	73
4.1	Diseño de clasificadores bajo enfoque máquina	73
4.1.1	Estimación paramétrica ML para clasificación	74
5	Filtrado Lineal	77
5.1	Introducción	77
5.2	El problema de filtrado	77
5.3	Solución ML	78
5.4	Filtro de Wiener	78
5.5	Solución Bayesiana	78
5.6	Cálculo online	79
5.6.1	Solución ML	79
5.6.2	Solución Bayesiana	80
5.7	Problemas	80
6	Soluciones de los problemas	81
6.1	Problemas del Capítulo 1	81
6.2	Problemas del Capítulo 3	82
6.3	Problemas del Capítulo 5	85
	References	87

Filtrado Lineal

5.1 Introduccion

Un problema común en estimación es el de querer determinar los coeficientes de un filtro lineal con M coeficientes a partir de la sola observación de las entradas y salidas de este. A esta tarea, así como a otras relacionadas, se la conoce con el nombre genérico de “filtrado lineal”. En este bloque mostraremos como las técnicas descritas en el bloque B1 pueden ser usadas para diseñar estimadores ML, MAP, MAD y MMSE de los coeficientes de dicho filtro, así como de futuras salidas del filtro si se conocen las correspondientes entradas.

5.2 El problema de filtrado

Suponga que se utiliza un filtro desconocido $\mathbf{s} = [s_1, s_2, \dots, s_M]^T$ para filtrar una señal $u[n]$. Al resultado se le suma cierto ruido gaussiano $\varepsilon[n]$ iid de media nula y varianza σ_ε^2 , dando lugar a una observación $x[n]$. Es decir,

$$x[n] = u[n]s_1 + u[n-1]s_2 + \dots + u[n-M+1]s_M + \varepsilon[n],$$

o, lo que es lo mismo, compactando $\mathbf{u}[n] = [u[n], u[n-1], \dots, u[n-M+1]]$, podemos decir que

$$x[n] = \mathbf{u}[n]\mathbf{s} + \varepsilon[n].$$

El problema de filtrado consiste entonces en estimar los coeficientes \mathbf{s} de un filtro a partir de un conjunto de entradas y salidas observadas, así como estimar la salida x_* correspondiente a una nueva entrada \mathbf{u}_* .

Si disponemos de las señales $u[n]$ y $x[n]$ en el intervalo $1 \leq n \leq N$, dispondremos de un total de $N - M + 1$ parejas entrada-salida, $\{\mathbf{u}[n], x[n]\}_{n=M}^N$. Estos datos observables pueden expresarse de forma más compacta como

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}[M] \\ \mathbf{u}[M+1] \\ \vdots \\ \mathbf{u}[N] \end{bmatrix} = \begin{bmatrix} u[M] & u[M-1] & \dots & u[1] \\ u[M+1] & u[M] & \dots & u[2] \\ \vdots & \vdots & \ddots & \vdots \\ u[N] & u[N-1] & \dots & u[N-M+1] \end{bmatrix}; \quad \mathbf{x} = \begin{bmatrix} x[M] \\ x[M+1] \\ \vdots \\ x[N] \end{bmatrix}.$$

5.3 Solución ML

El propio planteamiento del problema nos proporciona la verosimilitud de los coeficientes del filtro \mathbf{s} dada la observación n -ésima:

$$p(x[n]|\mathbf{s}) = \mathcal{N}(x[n]|\mathbf{u}[n]\mathbf{s}, \sigma_\varepsilon^2)$$

Dado un conjunto de observaciones, simplemente tomamos el producto de las anteriores verosimilitudes, ya que los términos de ruido son independientes

$$p(\mathbf{x}|\mathbf{s}) = \prod_{n=M}^N \mathcal{N}(x[n]|\mathbf{u}[n]\mathbf{s}, \sigma_\varepsilon^2) = \mathcal{N}(\mathbf{x}|\mathbf{U}\mathbf{s}, \sigma_\varepsilon^2\mathbf{I}).$$

El valor de \mathbf{s} que maximiza $p(\mathbf{x}|\mathbf{s})$ es

$$\hat{\mathbf{s}}_{\text{ML}} = (\mathbf{U}^\top \mathbf{U})^{-1} \mathbf{U}^\top \mathbf{x}$$

5.4 Filtro de Wiener

El filtro de Wiener $\mathbf{s}_{\text{Wiener}}$ es el filtro que minimiza el error cuadrático entre la salida observada $x[n]$ y la salida producida al ser utilizado para filtrar la entrada $u[n]$. Se puede demostrar fácilmente que dicho filtro puede expresarse como:

$$\mathbf{s}_{\text{Wiener}} = R_{uu}^{-1} r_{ux},$$

donde R_{uu} es la matriz de correlación de la señal de entrada $u[n]$ y r_{ux} es el vector de correlación cruzada entre $u[n]$ y $x[n]$. Desafortunadamente, estas dos cantidades son desconocidas en general, por lo que en la mayoría de las ocasiones, el filtro de Wiener no puede calcularse. Sin embargo, es frecuente usar la expresión anterior usando estimaciones muestrales para la matriz de correlación $\hat{R}_{uu} = \frac{1}{N-M+1} \mathbf{U}^\top \mathbf{U}$ y el vector de correlación cruzada $\hat{r}_{ux} = \frac{1}{N-M+1} \mathbf{U}^\top \mathbf{x}$. El resultado es una aproximación al filtro de Wiener $\hat{\mathbf{s}}_{\text{Wiener}} = \hat{R}_{uu}^{-1} \hat{r}_{ux}$ que minimiza el error cuadrático muestral (a menudo llamada “estimación least-squares”) y que coincide con la solución ML, es decir $\hat{\mathbf{s}}_{\text{Wiener}} = \hat{\mathbf{s}}_{\text{ML}}$.

A medida que el número de muestras disponibles para la estimación de los estadísticos R_{uu} y r_{ux} aumenta, dichas estimaciones se vuelven más precisas, de manera que $\hat{\mathbf{s}}_{\text{Wiener}}$ y por tanto $\hat{\mathbf{s}}_{\text{ML}}$ coinciden asintóticamente con el verdadero filtro de Wiener.

5.5 Solución Bayesiana

Para obtener un estimador Bayesiano de \mathbf{s} necesitamos conocer su probabilidad a priori $p(\mathbf{s})$. Aunque en general ésta es desconocida, es sensato utilizar

$$p(\mathbf{s}) = \mathcal{N}(\mathbf{s}|\mathbf{0}, \sigma_s^2\mathbf{I}),$$

ya que considera aceptable cualquier conjunto de coeficientes reales, y supone que estos tienen media nula y una dispersión fijada por σ_s^2 . También es posible fijar $\sigma_s^2 \rightarrow \infty$ para

conseguir una distribución uniforme. En cualquier caso, el uso de esta distribución a priori permite obtener de manera analítica la distribución a posteriori.

Conocidas la verosimilitud y la distribución a priori, podemos obtener la distribución a posteriori

$$p(\mathbf{s}|\mathbf{x}) = \mathcal{N}(\mathbf{s} \mid (\mathbf{U}^\top \mathbf{U} + \frac{\sigma_\varepsilon^2}{\sigma_s^2} \mathbf{I})^{-1} \mathbf{U}^\top \mathbf{x}, \sigma_\varepsilon^2 (\mathbf{U}^\top \mathbf{U} + \frac{\sigma_\varepsilon^2}{\sigma_s^2} \mathbf{I})^{-1}),$$

que inmediatamente nos proporciona los siguientes estimadores para \mathbf{s} :

$$\hat{\mathbf{s}}_{\text{MMSE}} = \hat{\mathbf{s}}_{\text{MAP}} = \hat{\mathbf{s}}_{\text{MAD}} = (\mathbf{U}^\top \mathbf{U} + \frac{\sigma_\varepsilon^2}{\sigma_s^2} \mathbf{I})^{-1} \mathbf{U}^\top \mathbf{x}$$

Nótese que suponer una distribución a priori uniforme (usando $\sigma_s^2 \rightarrow \infty$ en la expresión anterior) convierte la solución MAP en la solución ML obtenida anteriormente.

También podemos plantearnos el problema de estimar la salida x_* correspondiente a la entrada \mathbf{u}_* , a la vista de los mismos datos. Para ello, calculamos

$$\begin{aligned} p(x_*|\mathbf{x}) &= \int_{\mathbb{R}^M} p(x_*|\mathbf{s})p(\mathbf{s}|\mathbf{x})d\mathbf{s} \\ &= \mathcal{N}(x_* \mid \mathbf{u}_* (\mathbf{U}^\top \mathbf{U} + \frac{\sigma_\varepsilon^2}{\sigma_s^2} \mathbf{I})^{-1} \mathbf{U}^\top \mathbf{x}, \sigma_\varepsilon^2 + \sigma_\varepsilon^2 \mathbf{u}_* (\mathbf{U}^\top \mathbf{U} + \frac{\sigma_\varepsilon^2}{\sigma_s^2} \mathbf{I})^{-1} \mathbf{u}_*^\top), \end{aligned}$$

de manera que

$$\hat{x}_{*\text{MMSE}} = x_{*\text{MAP}} = x_{*\text{MAD}} = \mathbf{u}_* (\mathbf{U}^\top \mathbf{U} + \frac{\sigma_\varepsilon^2}{\sigma_s^2} \mathbf{I})^{-1} \mathbf{U}^\top \mathbf{x} = \mathbf{u}_* \hat{\mathbf{s}}_{\text{MMSE}}$$

5.6 Cálculo online

Es posible obtener las soluciones anteriores de manera online.

5.6.1 Solución ML

Se puede obtener una aproximación a $\hat{\mathbf{s}}_{\text{ML}}$ online con coste computacional $\mathcal{O}(M)$ sin más que notar que

$$\hat{\mathbf{s}}_{\text{ML}} = \underset{\mathbf{s}}{\operatorname{argmax}} p(\mathbf{x}|\mathbf{s}) = \underset{\mathbf{s}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{U}\mathbf{s}\|^2$$

y a continuación usar gradiente estocástico para minimizar $\|\mathbf{x} - \mathbf{U}\mathbf{s}\|^2$. La actualización de coeficientes que debe iterarse para efectuar la minimización es en este caso

$$\mathbf{s} \leftarrow \mathbf{s} - \mu(x[n] - \mathbf{u}[n]\mathbf{s})\mathbf{u}[n],$$

donde μ es un paso de adaptación “suficientemente pequeño”. A este algoritmo se le llama least mean squares (LMS).

5.6.2 Solución Bayesiana

Se puede obtener de manera exacta $\hat{\mathbf{s}}_{\text{MMSE}}$ a medida que se dispone de más muestras (N aumenta) sin necesidad de rehacer todos los cálculos, reusando la solución anterior. Para ello, se define $\mathbf{P}^{(N)} = (\mathbf{U}^\top \mathbf{U} + \frac{\sigma_s^2}{\sigma_e^2} \mathbf{I})^{-1}$, $\mathbf{r}^{(N)} = \mathbf{U}^\top \mathbf{x}$ y se usa el siguiente cálculo recursivo

$$\begin{aligned}\mathbf{P}^{(N+1)} &= \mathbf{P}^{(N)} - \frac{\mathbf{P}^{(N)} \mathbf{u}[N+1]^\top \mathbf{u}[N+1] \mathbf{P}^{(N)}}{1 + \mathbf{u}[N+1] \mathbf{P}^{(N)} \mathbf{u}[N+1]^\top} \\ \mathbf{r}^{(N+1)} &= \mathbf{r}^{(N)} + \mathbf{u}[N+1] x[N+1] \\ \mathbf{s}^{(N+1)} &= \mathbf{P}^{(N+1)} \mathbf{r}^{(N+1)},\end{aligned}$$

que sólo tiene un coste $\mathcal{O}(M^2)$ por paso (a diferencia de aplicar la ecuación original completa en cada paso, que tendría coste $\mathcal{O}(M^3)$). A este algoritmo se le llama recursive least squares (RLS).

5.7 Problemas

5.1. Considere la siguiente secuencia

$$u[1] \dots u[7] \equiv 0.7, -0.1, 0.7, -0.2, -0.1, 1.5, -1.1$$

que se alimenta como entrada a un filtro lineal de tres coeficientes $\mathbf{s} = [s_1, s_2, s_3]^\top$. Se conocen los siguientes elementos de la secuencia de salida, (corrompidos con ruido Gaussiano de varianza 0.25):

$$x[1] \dots x[6] \equiv -0.60, 1.13, 0.57, 0.42, 1.25, -2.58$$

- ¿Cuál es la estimación ML de \mathbf{s} ? (filtro de Wiener basado en estadísticos aproximados).
- Utilice el filtro obtenido para predecir $x[7]$, \hat{x}_{ML} .
- Calcule las estimación MMSE, MAP y MAD de \mathbf{s} asumiendo que la pdf a priori de sus componentes es $s_i \sim \mathcal{N}(0, 1)$.
- Obtenga la estimación MMSE de $x[7]$, \hat{x}_{MMSE} .
- Calcule el error cuadrático esperado en la predicción b). (Es decir, la esperanza de $(\hat{x}_{\text{ML}} - x[7])^2$ a la vista de los datos disponibles).
- Calcule el error cuadrático esperado en la predicción d). (Es decir, la esperanza de $(\hat{x}_{\text{MMSE}} - x[6])^2$ a la vista de los datos disponibles).

References

1. Hayes M H (1996) Statistical Digital Signal Processing and Modeling. John Wiley and Sons, New York, EE.UU.
2. Oppenheim A, Schaffer R (1999) Discrete-Time Signal Processing 2nd Ed. Prentice Hall, New York, EE.UU.
Thesis, Columbia University, New York