Mejor Predicción Lineal: inversión de un modelo lineal con ruido

Presentación vídeo: http://personales.upv.es/asala/YT/V/vcinv2.html

Objetivos: comprender la relación entre las fórmulas de "*mejor predicción lineal*" basadas en la matriz VC de dos variables aleatorias (a, b) y los modelos lineales con ruido $a = \theta b + \epsilon$ que se pueden identificar a partir de dichas fórmulas. En concreto, el objetivo es obtener el modelo "inverso", $b = \eta a + \varepsilon$ en un ejemplo (inverso en sentido "estadístico", de "mínima varianza del error").

Tabla de Contenidos

Preliminares, conceptos previos

Mejor predicción lineal

Dadas dos variables aleatorias a y b, suponemos media cero por simplificar (si no, se cambiaría todo a "incrementos sobre la media").

Si su matriz varianzas-covarianzas, simétrica, es:

$\Sigma := \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$, con $\Sigma_{ab} = \Sigma_{ba}^T$ (si fuera matriz, caso multivariable), entonces mejor pred. lineal de a dado b es:

$\widehat{a} = \Sigma_{ab}\Sigma_{bb}^{-1} \cdot b$, con una varianza del error de predicción $a - \widehat{a}$ dada por: $\Sigma_{e,a} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ab}^T$.

En caso de distribución normal, conocida varianza y media se conoce todo, con lo que la probabilidad condicional $a|b$ es una distribución $N(\widehat{a}, \Sigma_{e,a}) = N(\theta b, \Sigma_{e,a})$.

Matriz de varianzas-covarianzas asociada a un modelo lineal con ruido

Dado un modelo $a = \theta b + \epsilon$, con $b \sim N(0, \Sigma_{bb})$, $\epsilon \sim N(0, \Sigma_e)$, siendo ε independiente de b, entonces:

$$\Sigma_{aa} = E[(\theta b + \epsilon)(\theta b + \epsilon)^T] = E[\theta b b^T \theta^T + \epsilon b^T \theta^T + \theta b \epsilon^T + \epsilon \epsilon^T]$$
$$= \theta E[bb^T]\theta^T + E[\epsilon b^T]\theta^T + \theta E[b\epsilon^T] + E[\epsilon\epsilon^T] = \theta \Sigma_{bb} \theta^T + \Sigma_e$$

$$\Sigma_{ab} = E[(\theta b + \epsilon) b^T] = \theta E[bb^T] + E[\epsilon b^T] = \theta \Sigma_{bb}$$

Con lo que la matriz VC conjunta de a y b sería:

$$\Sigma := \begin{pmatrix} \theta \Sigma_{bb} \theta^T + \Sigma_e & \theta \Sigma_{bb} \\ \Sigma_{bb} \theta^T & \Sigma_{bb} \end{pmatrix}$$

El objetivo de este material es comprender la relación entre "mejor predicción lineal" y "modelo lineal con ruido" identificado.

Modelos identificados a partir de una matriz VC

Consideremos $\Sigma := \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$.

Si llamamos $\theta := \Sigma_{ab}\Sigma_{bb}^{-1}$, entonces la predicción es $\widehat{a} = \theta \cdot b$, con una varianza de error de predicción $\Sigma_{e,a} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1} \cdot \Sigma_{bb} \cdot \Sigma_{bb}^{-1}\Sigma_{ab}^T = \Sigma_{aa} - \theta \Sigma_{bb} \theta^T$.

Si ahora suponemos un modelo:

$a = \theta \cdot b + \epsilon$, siendo $\epsilon \sim N(0, \Sigma_{e,a})$ estadísticamente independiente de b, y b una variable aleatoria con media cero y varianza Σ_{bb}

entonces tendríamos $\Sigma_{aa} = E[(\theta b + \epsilon)(\theta b + \epsilon)^T] = \theta \Sigma_{bb} \theta^T + \Sigma_{e,a}$ que devuelve la expresión de arriba.

También $\Sigma_{ab} = \theta \Sigma_{bb} = \Sigma_{ab}\Sigma_{bb}^{-1} \cdot \Sigma_{bb}$ devuelve la covarianza inicial.

Por tanto, podríamos decir que el modelo lineal con ruido $a = \theta b + \epsilon$, con la varianza de b siendo Σ_{bb} y la varianza de ϵ siendo $\Sigma_{e,a}$ "explica" la matriz de varianzas covarianzas entre a y b.

Ejemplo: inversión de un modelo lineal (estático)

Supongamos que tenemos un modelo $b = coef \cdot a + \varepsilon$, con una varianza de a a priori de 4, varianza de ε de 1.75. El modelo inverso NO es $a = (b - \epsilon)/coef$, entendiendo "inverso" en el sentido estadístico de predicción "$\widehat{a}$" de "a" con la menor varianza posible del error $a - \widehat{a}$.

```
vza_a=4; vzaeps=1.75;
```

Entonces, la covarianza entre a y b es

```
coef=0.8;
covab=coef*vza_a
```

```
covab = 3.2000
```

y la varianza de b predicha por el modelo es

```
vza_b=vzaeps+coef*vza_a*coef
```

```
vza_b = 4.3100
```

```
MatrizVC_Sigma=[vza_a covab;covab vza_b]
```

```
MatrizVC_Sigma = 2×2
    4.0000    3.2000
    3.2000    4.3100
```

- La mejor predicción de b dado a es, obviamente, el modelo de partida:

```
covab/vza_a % = coef
```

```
ans = 0.8000
```

```
vzaerrb=vza_b-covab^2/vza_a % = vza eps
```

```
vzaerrb = 1.7500
```

- La mejor predicción de a dado b NO es $coef^{-1} \cdot b$:

```
eta=covab/vza_b
```

```
eta = 0.7425
```

```
1/coef
```

```
ans = 1.2500
```

```
vzaerra=vza_a-covab^2/vza_b
```

```
vzaerra = 1.6241
```

Otra forma de obtener la varianza del error al predecir "a dado b", $a - \eta b = [1 \quad -\eta] \cdot \begin{pmatrix} a \\ b \end{pmatrix}$

```
[1 -eta]*MatrizVC_Sigma*[1;-eta]
```

```
ans = 1.6241
```

- En efecto, si pruebo la varianza de $a - coef^{-1}b$, sale:

```
[1 -1/coef]*MatrizVC_Sigma*[1;-1/coef]
```

```
ans = 2.7344
```

y es más grande que `vzaerra`, por lo que el modelo "inverso algebraico" no es el de "mínima varianza del error de predicción".

Conclusiones

Invertir un modelo en estadística es más "complicado" que despejar de una ecuación algebraica determinista: añadir ruido hace que los datos "originales" sean irrecuperables.

En una "serie temporal" $x_{k+1} = \theta \cdot x_k + \epsilon_k$, algo parecido a esta ecuación se denomina ecuación "hacia adelante" (forward), y la ecuación "hacia atrás" (backwards) es $x_k = \eta x_{k+1} + \varepsilon'_{k+1}$, donde $\eta \neq \theta^{-1}$ y tampoco la varianza de ϵ_k y ε'_k coinciden con lo que la "inversión" $\theta^{-1}(x_k - \epsilon_{k-1})$ diría.