Fórmulas de predicción óptima lineal

Antonio Sala Piqueras

Dept. Ing. Sistemas y Automatica (DISA)
Universitat Politècnica de València (UPV)

*Presentación disponible en:

http:personales.upv.es/asala/YT/V/preli2.html

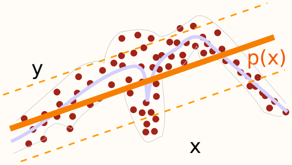
Presentación

Motivación:
La predicción óptima lineal (mínima varianza error, no correlado con información) puede obtenerse a partir de fórmulas relacionadas con las medias, varianzas y covarianzas de las variables implicadas.

Objetivos:
Comprender las relaciones entre las propiedades estadísticas de dos variables (o conjunto de ellas) y la fórmula de la mejor predicción lineal.

Contenidos:
Fórmula de mejor predicción (media). Calidad de la predicción (fórmula de varianza).

Predicción óptima LINEAL



Definición (Predicción óptima **lineal**)

La mejor predicción **lineal** de una variable aleatoria y dado un valor x_0 de otra variable aleatoria x es una función $p(x_0) = \Theta x_0 + \mu$ tal que el error de predicción $e = y - p(x)$ sea una variable aleatoria de **media cero**, **no correlada** con x (esto es, $\Sigma_{ex} = 0$), de **mínima varianza** entre todos los predictores **lineales** (variación total mínima, en caso multivariable).

Fórmula de predicción óptima lineal

Dadas dos variables aleatorias $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ de las que se conoce (analíticamente o experimentalmente –varianza y covarianza muestral–) medias $\mu_x = E(x)$, $\mu_y = E(y)$, y matriz de varianzas-covarianzas $(n+m) \times (n+m)$:

$$\Sigma = \begin{pmatrix} \Sigma_y & \Sigma_{yx} \\ \Sigma_{yx}^T & \Sigma_x \end{pmatrix}$$

Teorema

La predicción óptima lineal de y dada una muestra x_0 de x es:

$$p(x_0) = \boxed{(\Sigma_{yx}\Sigma_x^{-1})} \cdot (x_0 - \mu_x) + \mu_y$$

Prueba de la fórmula

(a) Media cero: $E[y - p(x)] = E[(y - \mu_y) - \Sigma_{yx}\Sigma_x^{-1}(x - \mu_x)] = 0$

(b) Error no correlado: $E[(y - p(x))(x - \mu_x)^T] =$
$E[((y - \mu_y) - \Sigma_{yx}\Sigma_x^{-1}(x - \mu_x))\,(x - \mu_x)^T] = \Sigma_{yx} - \Sigma_{yx}\Sigma_x^{-1}\Sigma_x = 0$

(c) Mínima varianza: en variables incrementales sobre la media,

$$E(ee^T) = E[(y - \Theta x)(y - \Theta x)^T] = E[yy^T + \Theta xx^T\Theta^T - yx^T\Theta^T - \Theta xy^T]$$
$$= \Sigma_y + \Theta\Sigma_x\Theta^T - \Sigma_{yx}\Theta^T - \Theta\Sigma_{yx}^T$$

Hagamos cambio de variable $\Theta = \Sigma_{yx}\Sigma_x^{-1} + \delta$. Entonces,

$$E(ee^T) = \Sigma_y + (\Sigma_{yx}\Sigma_x^{-1} + \delta)\Sigma_x(\Sigma_x^{-1}\Sigma_{yx}^T + \delta^T)$$
$$- \Sigma_{yx}(\Sigma_x^{-1}\Sigma_{yx}^T + \delta^T) - (\Sigma_{yx}\Sigma_x^{-1} + \delta)\Sigma_{yx}^T = \Sigma_y - \Sigma_{yx}\Sigma_x^{-1}\Sigma_{yx}^T + \delta\Sigma_x\delta^T$$

Como Σ_x es semidefinida positiva, $\delta\Sigma_x\delta^T$ también lo es (diagonal no negativa). Por lo tanto, sea cual sea δ la variación total del error (traza de $E[ee^T]$) aumenta (o se queda igual) comparado con $\delta = 0$.

Resumen:

[Ecuación de medias]
La mejor predicción lineal puntual de una variable aleatoria dada otra es:

$$p(x_0) = \boxed{(\Sigma_{yx}\Sigma_x^{-1})} \cdot (x_0 - \mu_x) + \mu_y$$

[Ecuación de varianzas]
El error de predicción $e = y - p(x)$ tiene una varianza dada por:

$$\boxed{\Sigma_e := E[e_p e_p^T] = \Sigma_y - \Sigma_{yx}\Sigma_x^{-1}\Sigma_{yx}^T}$$

Calidad del modelo óptimo lineal

$$\Sigma_e = \Sigma_y - \Sigma_{yx}\Sigma_x^{-1}\Sigma_{yx}^T$$

Teorema

Si dos variables x, y son **no correladas**, *los modelos lineales son inútiles para predecir una en función de la otra.*

*Si inicialmente x, y eran no correladas tenemos $\Sigma_{yx} = 0$, entonces $\Sigma_e = \Sigma_y$ (no mejoramos la calidad de la predicción de y conociendo x).

El teorema es una versión "lineal" del concepto de "estadísticamente independiente" (imposible predecir con **ningún** modelo).

Si $\Sigma_{yx} \neq 0$ entonces $\Sigma_e < \Sigma_y$, medir x sí **mejora** la calidad de la estimación de y.

Calidad del modelo óptimo lineal

$$\Sigma_e = \Sigma_y - \Sigma_{yx}\Sigma_x^{-1}\Sigma_{yx}^T$$

Teorema

Si dos variables x, y son **no correladas**, *los modelos lineales son inútiles para predecir una en función de la otra.*

*Si inicialmente x, y eran no correladas tenemos $\Sigma_{yx} = 0$, entonces $\Sigma_e = \Sigma_y$ (no mejoramos la calidad de la predicción de y conociendo x).

El teorema es una versión "lineal" del concepto de "estadísticamente independiente" (imposible predecir con **ningún** modelo).

Si $\Sigma_{yx} \neq 0$ entonces $\Sigma_e < \Sigma_y$, medir x sí **mejora** la calidad de la estimación de y.

Conclusiones

- La predicción óptimal lineal puede obtenerse con una fórmula a partir de la matriz de varianzas-covarianzas de los datos intervinientes: error de predicción de mínima varianza, no correlado con datos de entrada, (subóptimo si la relación "real" es no lineal).
- Coincide con el estimador "determinista" de mínimos cuadrados (no demostrado aquí).
- Si existe correlación entre dos variables, el error **a posteriori** ($\Sigma_y - \Sigma_{yx}\Sigma_x^{-1}\Sigma_{yx}^T$, utilizando la mejor predicción) es **más pequeño** que el error **a priori** (Σ_y antes de usar la información).