Normalización de las variables

Supóngase una variable que esté definida en un intervalo:

$x = [x_{min}, x_{max}]$ (1)

El objetivo es normalizar la variable, lo cual significa que debe convertirse a un nuevo intervalo de definición:

$x_n = [-n_{norm}, n_{norm}]$ (2)

Para resolver este problema se puede definir el siguiente sistema de ecuaciones lineales:

$kx_{min} + p = -n_{norm}$ (3)

$kx_{max} + p = n_{norm}$ (4)

donde k y p se denominan *factores de escala*. La solución al anterior sistema de ecuaciones es:

$k = \frac{2n_{norm}}{x_{max} - x_{min}}$ (5)

$p = -\left[\frac{x_{max} + x_{min}}{x_{min} - x_{max}}\right] n_{norm}$ (6)

Las anteriores expresiones no serían válidas si en (1) ambos valores son negativos.

La forma más común de normalizar es definir (2) como:

$x_n = [-1, 1]$ o $n_{norm} = 1$ (7)

por lo que (5) y (6) resulta:

$k = \frac{2}{x_{max} - x_{min}}$ (8)

$p = -\left[\frac{x_{max} + x_{min}}{x_{max} - x_{min}}\right]$ (9)

Ejemplo 1:

Supóngase un intervalo (1) definido por:

$x = [4, 25]$

se define la forma (7). De aplicar (8) y (9) se tiene:

$k = \frac{2}{21} = 0.095; \quad p = -\frac{29}{21} = -1.38$

Para comprobar, el resultado se sustituye en (3) y (4), de lo que resulta:

$0.095(4) - 1.38 = -1$

$0.095(25) - 1.38 = 0.995 \cong 1$

Ejemplo 2:

Supóngase un intervalo (1) definido por:

$x = [-3, 67]$

se define la forma (7). De aplicar (8) y (9) se tiene:

$k = \frac{2}{70} = 0.028; \quad p = -\frac{64}{70} = -0.914$

Para comprobar, el resultado se sustituye en (3) y (4), de lo que resulta:

$0.028(-3) - 0.914 = -0.998 \cong -1$

$0.028(67) - 0.914 = 0.962 \cong 1$

Nótese en el ejemplo anterior como la aproximación realizada en los cálculos de k y p tienen una mayor influencia en la precisión de la normalización de los datos en la medida en que el intervalo que se pretende normalizar es mayor.

De (3) y (4) se deduce que para normalizar cualquier valor $x \in [x_{min}, x_{max}]$ se debe aplicar:

$n_{norm} = kx + p$ (10)

donde n_{norm} representa el número normalizado.

Se pueden entonces definir los siguientes pasos para normalizar las variables al intervalo [-1, 1], a partir de dos posibles casos:

(a) $[x_{min}, x_{max}] \equiv [+, +]$ ó $[x_{min}, x_{max}] \equiv [-, +]$:
Se calculan k y p utilizando (8) y (9).
El valor normalizado se obtiene aplicando la expresión (10).

(b) $[x_{min}, x_{max}] \equiv [-, -]$:
Se convierten a $[x_{min}, x_{max}] \equiv [+, +]$ y se aplican los pasos definidos en (a).

En el caso en que se normalice la variable de salida y se obtenga un modelo, entonces será necesario desnormalizar el valor de salida del núcleo estimador, para ello se aplica (de (10))

$x = \frac{n_{norm} - p}{k}$ (11)