Clasificación Supervisada

Funciones Discriminantes

Jesús Ariel Carrasco Ochoa
Instituto Nacional de Astrofísica, Óptica y Electrónica

Funciones Discriminantes

- Sea T un conjunto de entrenamiento con n objetos $O_1,\ldots,O_n$
- Los objetos de T están agrupados en 2 clases C_1 y C_2 y se conoce a la clase de cada objeto
- Sea n_i el número de objetos de T en C_i
- Cada objeto O_j descrito por m variables $x_1,\ldots,x_m$

Funciones Discriminantes

Encontrar una función ***g(X)*** que genere valores positivos para los objetos de $\boldsymbol{C_1}$ y negativos para los objetos de $\boldsymbol{C_2}$

Dado un nuevo objeto ***O***

$$\text{Si} \quad g(O) \geq 0 \;\rightarrow\; C_O = C_1$$

$$\text{Si} \quad g(O) < 0 \;\rightarrow\; C_O = C_2$$

Funciones Discriminantes

- Sea $\boldsymbol{T}$ un conjunto de entrenamiento con $\boldsymbol{n}$ objetos $\boldsymbol{O_1,\dots,O_n}$
- Los objetos de $\boldsymbol{T}$ están agrupados en r clases $\boldsymbol{C_1,\dots,C_r}$ y se conoce a la clase de cada objeto
- Sea $\boldsymbol{n_i}$ el número de objetos de $\boldsymbol{T}$ en $\boldsymbol{C_i}$
- Cada objeto $\boldsymbol{O_j}$ descrito por $\boldsymbol{m}$ variables $\boldsymbol{x_1,\dots,x_m}$
- Se tiene una función $\boldsymbol{d}$ que evalúa la distancia entre objetos

Funciones Discriminantes

Encontrar una función $\boldsymbol{g_i(X)}$ para cada clase, tal que dado un nuevo objeto ***O***, éste se ubica en $C_{\boldsymbol{o}}$ si

$$C_O = \max_{C_i} \{ g_i(O) \}$$

Funciones Discriminantes

Anteriormente teníamos una ***f(X)*** tal que

$$C_O = \min_{C_i}\left\{f_i(O)\right\} \quad \text{o} \quad C_O = \max_{C_i}\left\{f_i(O)\right\}$$

Con lo cual para 2 clases podemos construir ***g(X)*** como

$$g(O) = f_2(O) - f_1(O) \quad \text{o} \quad g(O) = f_1(O) - f_2(O)$$

Funciones Discriminantes

Si ***g(X)*** es lineal debe tener la forma

$$g(O) = w_0 + W^T X(O)$$

$$g(O) = w_0 + \sum_{j=1}^{m} w_j x_j(O)$$

Funciones Discriminantes

Con el clasificador basado en la distancia Euclidiana a la media de la clase tenemos

$$
\begin{aligned}
g(O) &= \left\| X(O) - \mu_2 \right\|^2 - \left\| X(O) - \mu_1 \right\|^2 \\
&= (X(O) - \mu_2)^T (X(O) - \mu_2) - (X(O) - \mu_1)^T (X(O) - \mu_1) \\
&= \quad \left(X(O)^T X(O) - 2\mu_2{}^T X(O) + \mu_2{}^T \mu_2 \right) \\
&\quad - \left(X(O)^T X(O) - 2\mu_1{}^T X(O) + \mu_1{}^T \mu_1 \right) \\
&= -2\mu_2{}^T X(O) + \mu_2{}^T \mu_2 + 2\mu_1{}^T X(O) - \mu_1{}^T \mu_1 \\
&= \left(\mu_2{}^T \mu_2 - \mu_1{}^T \mu_1 \right) + 2\left(\mu_2 - \mu_1 \right)^T X(O)
\end{aligned}
$$

Funciones Discriminantes

Con el clasificador basado en la distancia Euclidiana a la media de la clase tenemos

$$g(O) = \left({\mu_2}^T \mu_2 - {\mu_1}^T \mu_1 \right) + 2\left(\mu_2 - \mu_1\right)^T X(O)$$

Con esto tenemos

$$w_0 = \left({\mu_2}^T \mu_2 - {\mu_1}^T \mu_1 \right)$$
$$W = 2\left(\mu_2 - \mu_1\right)^T$$

Funciones Discriminantes

Clasificador lineal de Fisher

Se quiere maximizar: $$\frac{D^2}{SS}$$

donde

$$D = (\mu_1 - \mu_2)^t W$$

$$SS = \sum_{j=1}^{m} \sum_{j=1}^{m} w_j w_k \Sigma_{jk}$$

Funciones Discriminantes

Clasificador lineal de Fisher

$$w_j = \sum_{k=1}^{m} \Sigma_{jk}^{-1} \left(\mu_{1j} - \mu_{2j} \right) \quad \text{para } j = 1, \ldots, m$$

$$w_0 = -\sum_{j=1}^{m} w_j \frac{\left(\mu_{1j} + \mu_{2j} \right)}{2}$$

Con esto tenemos

$$g(O) = w_0 + W^T X(O)$$

Funciones Discriminantes

El clasificador lineal de Fisher es equivalente al clasificador basado en la distancia de Mahalanobis a la media de la clase

$$g(O) = D_M(X(O) - \mu_2)^2 - D_M(X(O) - \mu_1)^2$$

donde D_M es la distancia de Mahalanobis

Funciones Discriminantes

Si tomamos

$$Y = \left(1, x_1(O), \ldots, x_m(O)\right)$$

$$W = (w_0, w_1, \ldots, w_m)$$

Tenemos que

$$g(O) = w_0 + \sum_{j=1}^{m} w_j x_j(O) = WY$$

Funciones Discriminantes

Método de incremento fijo (2 clases)

1) Construir $\boldsymbol{Y_j}$ para cada $\boldsymbol{O_j \in T}$ como

$$Y_j = \left(1, x_1(O_j), \ldots, x_m(O_j)\right)$$

2) Inicializar $\boldsymbol{W}=(\boldsymbol{w_0}, \boldsymbol{w_1}, \ldots, \boldsymbol{w_m})$ con valores aleatorios

Funciones Discriminantes

Método de incremento fijo (2 clases)

3) Si con $\boldsymbol{W}$ se pueden clasificar bien todos los objetos de $\boldsymbol{T}$ → FIN

4) Tomar un objeto $\boldsymbol{O_j}$ que sea mal clasificado con $\boldsymbol{W}$

Funciones Discriminantes

Método de incremento fijo (2 clases)

5) Si $\boldsymbol{O_j \in C_1}$ tomar $W = W + Y_j$

6) Si $\boldsymbol{O_j \in C_2}$ tomar $W = W - Y_j$

7) Ir a paso 3

Funciones Discriminantes

Método de incremento fijo (2 clases)

Si $\boldsymbol{C_1}$ y $\boldsymbol{C_2}$ son linealmente separables → $\boldsymbol{W}$ converge a la recta que las separa

El paso 7 debe limitarse con un número máximo de iteraciones

Funciones Discriminantes

Método de incremento fijo (2 clases)

Por qué? si $\boldsymbol{O_j \in C_1}$ y fue mal clasificado

$$W = W + Y_j$$

Porque se quiere que $\boldsymbol{WY_j} \geq 0$ pero como $\boldsymbol{O_j}$ fue mal clasificado $\boldsymbol{WY_j} < 0$

$$(W + Y_j)Y_j = WY_j + Y_jY_j$$

Funciones Discriminantes

Método de incremento fijo (2 clases)

Como $\boldsymbol{Y_{j1}}=1 \rightarrow \boldsymbol{Y_j Y_j}>0$

$$(W+Y_j)Y_j > WY_j$$

Entonces $\boldsymbol{WY_j}$ irá creciendo hasta que finalmente sea ≥ 0

Funciones Discriminantes

Método de incremento fijo (2 clases)

Por qué? si $\boldsymbol{O_j \in C_2}$ y fue mal clasificado

$$W = W - Y_j$$

Porque se quiere que $\boldsymbol{WY_j}<0$ pero como $\boldsymbol{O_j}$ fue mal clasificado $\boldsymbol{WY_j}\geq 0$

$$(W - Y_j)Y_j = WY_j - Y_jY_j$$

Funciones Discriminantes

Método de incremento fijo (2 clases)

Como $\boldsymbol{Y_{j1}}=1 \rightarrow \boldsymbol{Y_j Y_j}>0$

$$(W - Y_j)Y_j < WY_j$$

Entonces $\boldsymbol{WY_j}$ irá decreciendo hasta que finalmente sea <0

Funciones Discriminantes

Método de incremento fijo (***r*** clases)

1) Construir $\boldsymbol{Y_j}$ para cada $\boldsymbol{O_j \in T}$ como

$$Y_j = \left(1, x_1(O_j), \ldots, x_m(O_j)\right)$$

2) Inicializar $\boldsymbol{W_i}=(\boldsymbol{w_{i0}, w_{i1}, \ldots, w_{im}})$ con valores aleatorios, $\boldsymbol{i}=1,\ldots,\boldsymbol{r}$

Funciones Discriminantes

Método de incremento fijo ($\boldsymbol{r}$ clases)

3) Si con $\boldsymbol{W_1,\ldots,W_r}$ se pueden clasificar bien todos los objetos de $\boldsymbol{T} \rightarrow$ FIN

4) Tomar un objeto $\boldsymbol{O_j}$ que sea mal clasificado

Funciones Discriminantes

Método de incremento fijo (***r*** clases)

5) Si $\boldsymbol{O_j \in C_i}$ tomar

$$W_i = W_i + Y_j$$

$$\left(\forall k \middle| g_i(O_j) \leq g_k(O_j)\right) W_k = W_k - Y_j$$

6) Ir a paso 3

Funciones Discriminantes

Método del Gradiente (2 clases)

1) Construir $\boldsymbol{Y_j}$ para cada $\boldsymbol{O_j} \in \boldsymbol{T}$ como

$$Y_j = \left(1, x_1(O_j), \ldots, x_m(O_j)\right)$$

2) Para todo $\boldsymbol{O_j} \in \boldsymbol{C_2}$ tomar $\boldsymbol{Y_j}=(-1)\boldsymbol{Y_j}$

3) Inicializar $\boldsymbol{W}=(\boldsymbol{w_0}, \boldsymbol{w_1}, \ldots, \boldsymbol{w_m})$ con valores aleatorios

Funciones Discriminantes

Método del Gradiente (2 clases)

Nota: Un objeto será bien clasificado si $\boldsymbol{WY_j} \geq 0$

Tomar

$$H = -\sum_{j=1}^{n} h_j$$

$$h_j = \begin{cases} 0 & \text{si } WY_j \geq 0 \\ -WY_j & \text{si } WY_j < 0 \end{cases}$$

Funciones Discriminantes

Método del Gradiente (2 clases)

4) Si $\boldsymbol{H}=0 \rightarrow$ FIN

5) Sean $\boldsymbol{O}_{g1},\ldots,\boldsymbol{O}_{gs}$ los objetos mal clasificados con $\boldsymbol{W}$

6) Tomar

$$W = W + \sum_{j=1}^{s} Y_{gj}$$

7) Ir a paso 4

Funciones Discriminantes

Método del Gradiente (2 clases)

Si $\boldsymbol{C_1}$ y $\boldsymbol{C_2}$ son linealmente separables → $\boldsymbol{W}$ converge a la recta que las separa

El paso 7 debe limitarse con un número máximo de iteraciones

Clasificación Supervisada

Funciones Discriminantes

Jesús Ariel Carrasco Ochoa
Instituto Nacional de Astrofísica, Óptica y Electrónica