Tarea 2: Clasificador bayesiano ingenuo

Saul Ivan Rivas Vega

Aprendizaje Automatizado

5 de marzo de 2020

1. Géneros

Un programa de salud gubernamental desea clasificar los registros de las personas en géneros femenino (F) o masculino (M) a partir de los atributos nombre, estatura y peso. Se cuentan con los siguientes registros:

Nombre	Estatura (m)	Peso (Kg)	Género
Denis	1.72	75.3	M
Guadalupe	1.82	81.6	M
Alex	1.80	86.1	M
Alex	1.70	77.1	M
Cris	1.73	78.2	M
Juan	1.80	74.8	M
Juan	1.80	74.3	M
Denis	1.50	50.5	F
Alex	1.52	45.3	F
Cris	1.62	61.2	F
Rene	1.67	68.0	F
Guadalupe	1.65	58.9	F
Guadalupe	1.75	68.0	F

Entrena un clasificador bayesiano ingenuo usando estimación por máxima verosimilitud y otro usando estimación por máximo a posteriori. Reporta los parámetros que obtuviste en ambos casos y usa los clasificadores entrenados para predecir la clase de los siguientes vectores: x1 = (Rene, 1.68, 65), x2 = (Guadalupe, 1.75, 80), x3 = (Denis, 1.80, 79), x4 = (Alex, 190, 85) y x5 = (Cris, 165, 70). Describe de forma detallada el procedimiento que seguiste tanto en el entrenamiento como en la predicción y discute los resultados obtenidos. Para el entrenamiento del clasificador por máximo a posteriori considera los siguientes valores para las distribuciones correspondientes:

Género	Nombre	Estatura			Peso		
	α_k	μ_0	σ_0^2	σ^2	μ_0	σ_0^2	σ^2
M	$1, \forall k$	1.7	0.3	0.0020	85.5	17.0	15.76
F	$1, \forall k$	1.5	0.1	0.0074	70.3	85.0	71.00

1.1. Estimador por máxima verosimilitud

Los atributos son: **nombre**, **estatura** y **peso**, y la clase es **género**.

1.1.1. Atributo ***Nombre***

Para el **nombre** podemos asumir una distribución categórica:

$$X_{nombre}^{(i)} \sim Cat(X_{nombre}^{(i)}; q) \tag{1}$$

Donde las categorías son los nombres y los podemos enumerar:

1. Denis
2. Guadalupe
3. Alex
4. Cris
5. Juan
6. Rene

Y así con los nombres de 1 a 6 podemos definir a $Cat(X_{nombre}^{(i)}; q)$ como:

$$Cat(X_{nombre}^{(i)}; q) = \prod_{k=1}^{6} q_k^{[x_{nombre}^{(i)}=k]} \tag{2}$$

Donde podemos estimar a q_k usando el estimador de máxima verosimilitud como:

$$\begin{gathered} \hat{q}_k = \frac{c_k}{n} \\ \text{Donde } c_k\text{:} \\ c_k = \sum_{i=1}^{n} [x_{nombre}^{(i)} = k] \end{gathered} \tag{3}$$

Así podemos estimar el parámetro para la primer categoría:

Para la clase **Femenino** :

$$
\begin{aligned}
c_{1F} &= \sum_{i=1}^{13} [x^{(i)}_{nombre} = 1 \text{ y } x^{(i)} \text{ es de la clase Femenino}] \\
&= 0+0+0+0+0+0+0+0+1+0+0+0+0+0 \\
&= 1 \\
\hat{q}_{(1|F)} =& \frac{1}{6}
\end{aligned} \tag{4}
$$

Para la clase **Masculino** :

$$
\begin{aligned}
c_{1M} &= \sum_{i=1}^{13} [x^{(i)}_{nombre} = 1 \text{ y } x^{(i)} \text{ es de la clase Masculino}] \\
&= 1+0+0+0+0+0+0+0+0+0+0+0+0+0 \\
&= 1 \\
\hat{q}_{(1|M)} =& \frac{1}{7}
\end{aligned}
$$

Y de la misma forma para las categorías restantes:

$$
\begin{aligned}
&c_{2F} = 2,\ \hat{q}_{(2|F)} = \frac{2}{6},\ c_{2M} = 1,\ \hat{q}_{(2|M)} = \frac{1}{7} && c_{3F} = 1,\ \hat{q}_{(3|F)} = \frac{1}{6},\ c_{3M} = 2,\ \hat{q}_{(3|M)} = \frac{2}{7} \\
&c_{4F} = 1,\ \hat{q}_{(4|F)} = \frac{1}{6},\ c_{4M} = 1,\ \hat{q}_{(4|M)} = \frac{1}{7} && c_{5F} = 0,\ \hat{q}_{(5|F)} = \frac{0}{6} = 0,\ c_{5M} = 2,\ \hat{q}_{(5|M)} = \frac{2}{7} \\
&c_{6F} = 1,\ \hat{q}_{(6|F)} = \frac{1}{6},\ c_{6M} = 0,\ \hat{q}_{(6|M)} = \frac{0}{7} = 0
\end{aligned} \tag{5}
$$

1.1.2. Atributo Estatura

Para la **estatura** podemos asumir una distribución normal:

$$
X^{(i)}_{estatura} \sim \mathcal{N}(X^{(i)}_{estatura}; \mu; \sigma^2) \tag{6}
$$

Donde $\mathcal{N}(X^{(i)}_{estatura}; \mu; \sigma^2)$ se define como:

$$
\mathcal{N}(X^{(i)}_{estatura}; \mu; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x^{(i)}-\mu)^2}{2\sigma^2}} \tag{7}
$$

Donde podemos estimar a μ y a σ usando el estimador de máxima verosimilitud como:

Para la clase **Femenino**:

$$\begin{aligned}
\hat{\mu}_F &= \frac{1}{n}\sum_{i=1}^{n} x_{estatura}^{(i)} \\
&= \frac{1}{6}(1.50 + 1.52 + 1.62 + 1.67 + 1.65 + 1.75) \\
&= \frac{1}{6}(9.71) \\
&= 1.618\bar{3}
\end{aligned} \tag{8}$$

$$\begin{aligned}
\hat{\sigma}_F^2 &= \frac{1}{n}\sum_{i=1}^{n} (x_{estatura}^{(i)} - \hat{\mu})^2 \\
&= \frac{1}{6}[(1.50 - 1.618\bar{3})^2 + (1.52 - 1.618\bar{3})^2 + \cdots + (1.75 - 1.618\bar{3})^2)] \\
&= \frac{1}{6}[0.00046 + 0.01477 + 0.01031 + \cdots + 0.00266] \\
&= \frac{1}{6}[0.13036923076923082] \\
&= 0.021728205128205138
\end{aligned} \tag{9}$$

1.1.3. Atributo Peso

Para el **peso** podemos asumir una distribución normal:

$$X_{peso}^{(i)} \sim \mathcal{N}(X_{peso}^{(i)}; \mu; \sigma^2) \tag{10}$$

Donde $\mathcal{N}(X_{peso}^{(i)}; \mu; \sigma^2)$ se define como:

$$\mathcal{N}(X_{peso}^{(i)}; \mu; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x^{(i)}-\mu)^2}{2\sigma^2}} \tag{11}$$

Donde podemos estimar a μ y a σ usando el estimador de máxima verosimilitud como:

$$\begin{aligned}
\hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} x_{peso}^{(i)} \\
&= \frac{1}{13}(75.3 + 81.6 + 86.1 + 77.1 + 78.2 + 74.8 + 74.3 + 50.5 + 45.3 + 61.2 + 68.0 + 58.9 + 68.0) \\
&= \frac{1}{13}(899.3) \\
&= 69.17692307692307
\end{aligned} \tag{12}$$

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_{peso}^{(i)} - \hat{\mu})^2 \\
&= \frac{1}{13}[(75.3 - 69.17692)^2 + (81.6 - 69.17692)^2 + \cdots + (68.0 - 69.17692)^2)] \\
&= \frac{1}{13}[37.49207 + 154.33284 + 286.39053 + \cdots + 1.38514] \\
&= \frac{1}{13}[1771.2230769230764] \\
&= 136.2479289940828
\end{aligned} \tag{13}$$

1.1.4. Género

2. Spam