

Aprendizaje Basado en Instancias

García Prado, Sergio
sergio@garciparedes.me

8 de abril de 2017

Resumen

[TODO]

1. INTRODUCCIÓN

[TODO]

2. LA FIGURA 1 MUESTRA UN CONJUNTO DE ENTRENAMIENTO CON EJEMPLOS POSITIVOS (ESTRELLAS) Y NEGATIVOS (CÍRCULOS). SE DESEA CLASIFICAR LA NUEVA INSTANCIA $\langle 3, 3 \rangle$ MEDIANTE EL ALGORITMO K -VECINOS MÁS PRÓXIMOS. OBTENER LA CLASIFICACIÓN PARA LOS VALORES DE $K = \{1, 3, 5\}$ UTILIZANDO LAS DISTANCIAS INDICADAS A CONTINUACIÓN

[TODO]

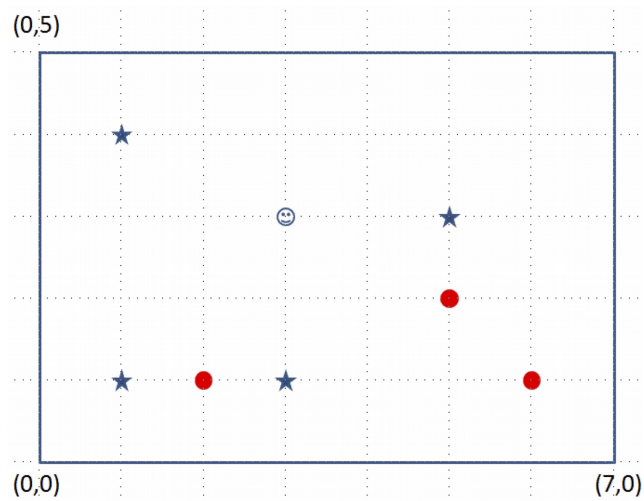


Figura 1: Representación Gráfica del problema 2

$$ESTRELLAS = \{(1, 1), (1, 4), (3, 1), (5, 3)\} \quad (1)$$

$$CIRCULOS = \{(2, 1), (5, 2), (6, 1)\} \quad (2)$$

$$instancia = (3, 3) \quad (3)$$

$$max_x = 6 \ min_x = 1 \ max_y = 4 \ min_y = 1$$

$$normalize(x, y) = (\frac{x - min_x}{max_x - min_x}, \frac{y - min_y}{max_y - min_y}) \quad (4)$$

$$ESTRELLAS_{normalized} = \{(0, 0), (0, 1), (\frac{2}{5}, 0), (\frac{4}{5}, \frac{2}{3})\} \quad (5)$$

$$CIRCULOS_{normalized} = \{(\frac{1}{5}, 0), (\frac{4}{5}, \frac{1}{3}), (1, 0)\} \quad (6)$$

$$instancia_{normalized} = (\frac{2}{5}, \frac{2}{3}) \quad (7)$$

2.1. DISTANCIA EUCLÍDEA

$$D_{euclidean}(a, b) = \sqrt{(a_x - b_x)^2 + (a_y - b_y)^2} \quad (8)$$

$$R_{euclidean} = \{4.242_e, 3.605_e, 3.969_e, 3.206_e, 4.103_c, 3.457_c, 3.605_c\} \quad (9)$$

[TODO]

$$min(R_{euclidean}, 1) = \{3.206_e\} \quad (10)$$

$$min(R_{euclidean}, 3) = \{3.206_e, 3.457_c, 3.605_e\} \quad (11)$$

Empate

$$min(R_{euclidean}, 5) = \{3.206_e, 3.457_c, 3.605_e, 3.605_c, 3.969_e\} \quad (12)$$

2.2. DISTANCIA EUCLÍDEA PONDERADA: $w_x = 0.2, w_y = 0.8$

[TODO]

$$D_{w_{euclidean}}(a, b) = \sqrt{w_x(a_x - b_x)^2 + w_y(a_y - b_y)^2} = \sqrt{0.2(a_x - b_x)^2 + 0.8(a_y - b_y)^2} \quad (13)$$

$$R_{w_{euclidean}} = \{3.0_e, 2.236_e, 2.924_e, 2.307_e, 2.961_c, 2.580_c, 2.828_c\} \quad (14)$$

$$min(R_{w_{euclidean}}, 1) = \{2.236_e\} \quad (15)$$

$$min(R_{w_{euclidean}}, 3) = \{2.236_e, 2.307_e, 2.580_c\} \quad (16)$$

$$min(R_{w_{euclidean}}, 5) = \{2.236_e, 2.307_e, 2.580_c, 2.828_c, 2.924_e\} \quad (17)$$

2.3. DISTANCIA DE MANHATTAN

[TODO]

$$D_{manhattan}(a, b) = |a_x - b_x| + |a_y - b_y| \quad (18)$$

$$R_{manhattan} = \{4_e, 3_e, 2_e, 2_e, 3_c, 3_c, 5_c\} \quad (19)$$

$$\min(R_{manhattan}, 1) = \{2_e\} \quad (20)$$

$$\min(R_{manhattan}, 3) = \{2_e, 2_e, 3_e\} \quad (21)$$

$$\min(R_{manhattan}, 5) = \{2_e, 2_e, 3_e, 3_c, 3_c, \} \quad (22)$$

2.4. DISTANCIA DE HAMMING

[TODO]

$$D_{hamming}(a, b) = (a_x \neq b_x) + (a_y \neq b_y) \quad (23)$$

$$R_{hamming} = \{2_e, 2_e, 1_e, 1_e, 2_c, 2_c, 2_c\} \quad (24)$$

$$\min(R_{manhattan}, 1) = \{1_e\} \quad (25)$$

$$\min(R_{manhattan}, 3) = \{1_e, 1_e, 2_e\} \quad (26)$$

$$\min(R_{manhattan}, 5) = \{1_e, 1_e, 2_e, 3_c, 3_c, \} \quad (27)$$

3. DÍGITOS MANUSCRITOS

[TODO]

Validación cruzada de 10 particiones — K -Vecinos más Próximos									
Datos	Tasa de Error ($K =$)								
	1	2	3	4	5	6	7	8	9
Entrenae	3.448 %	3.524 %	3.065 %	3.141 %	3.371 %	3.218 %	3.295 %	3.295 %	3.371 %

Tabla 1: Tasa de error obtenida tras realizar un experimento de Validación cruzada de 10 particiones con el clasificador K-NN para $k \in \{1, 2, \dots, 9\}$

[TODO]

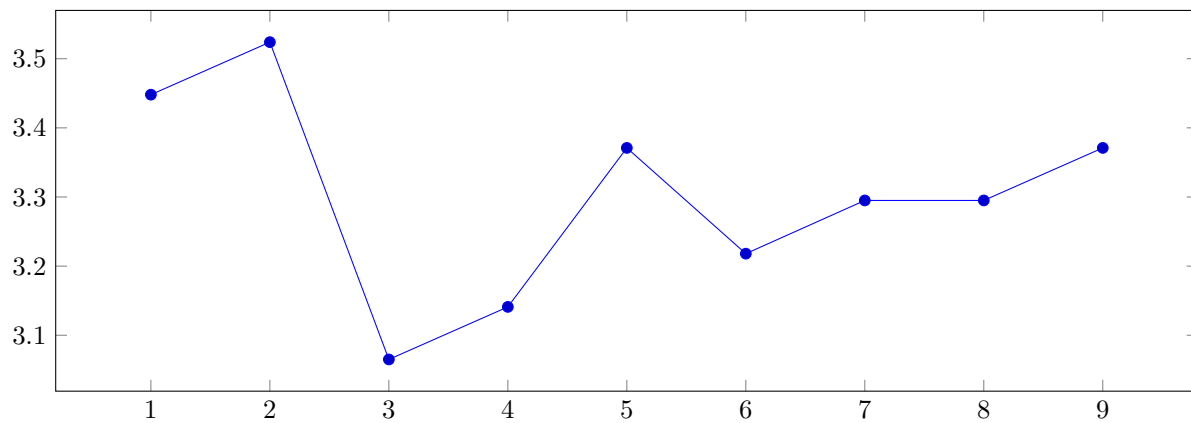


Figura 2: Representación Gráfica de la tasa de error obtenida tras realizar un experimento de Validación cruzada de 10 particiones con el clasificador K-NN para $k \in \{1, 2, \dots, 9\}$

REFERENCIAS

- [CCAG17] Teodoro Calonge Cano and Carlos Javier Alonso González. Técnicas de Aprendizaje Automático, 2016/17.
- [GP17] Sergio García Prado. Aprendizaje basado en instancias. <https://github.com/garciparedes/machine-learning-instance-based>, 2017.
- [too] Weka. <http://www.cs.waikato.ac.nz/ml/weka/>.