Database Systems Group • Prof. Dr. Thomas Seidl

Exercise 10: SVM + Decision Tree

Knowledge Discovery in Databases I

SS 2016

Exercise 10-1: SVM

$$x_1 = (2,3), x_2 = (3,2), x_3 = (4,4), x_4 = (4,2)$$

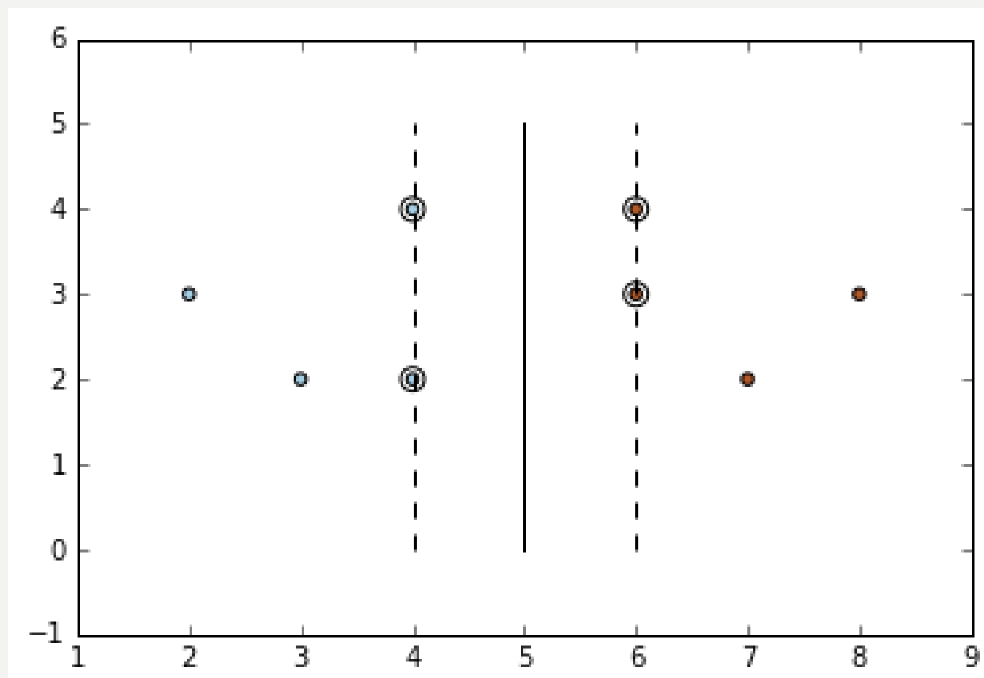
$$x_5 = (6,4), x_6 = (6,3), x_7 = (7,2), x_8 = (8,3)$$

$$A = \{x_1, x_2, x_3, x_4\}, B = \{x_5, x_6, x_7, x_8\}$$

$$x_1 = (2,3), x_2 = (3,2), x_3 = (4,4), x_4 = (4,2)$$

$$x_5 = (6,4), x_6 = (6,3), x_7 = (7,2), x_8 = (8,3)$$

$$A = \{x_1, x_2, x_3, x_4\}, B = \{x_5, x_6, x_7, x_8\}$$

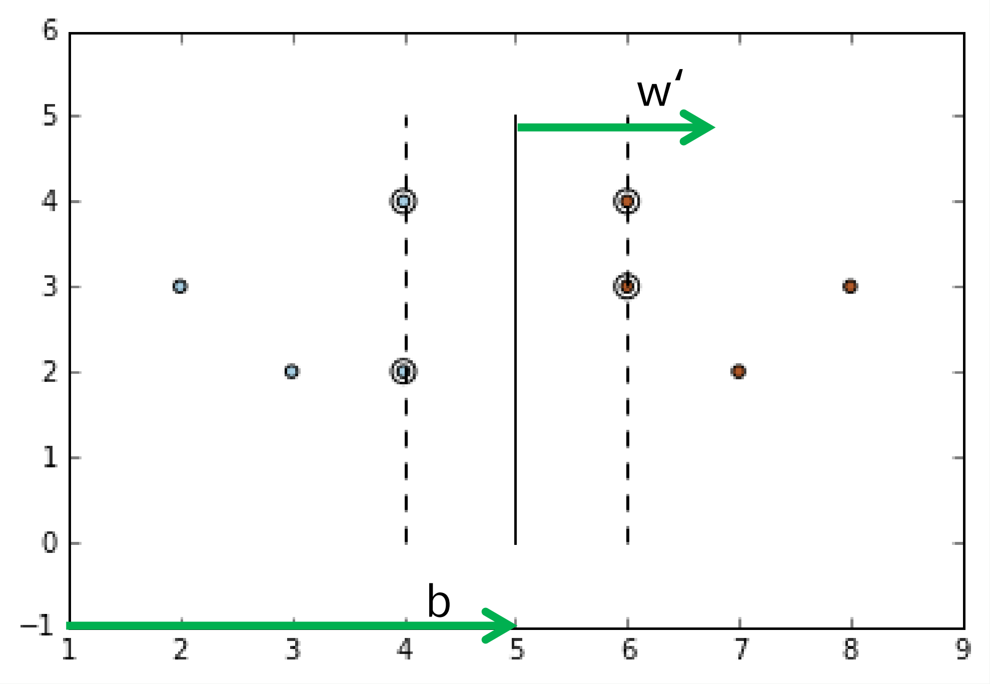


Exercise 10-1: SVM

$x_1 = (2, 3), x_2 = (3, 2), x_3 = (4, 4), x_4 = (4, 2)$

$x_5 = (6, 4), x_6 = (6, 3), x_7 = (7, 2), x_8 = (8, 3)$

$A = \{x_1, x_2, x_3, x_4\}, B = \{x_5, x_6, x_7, x_8\}$



Exercise 10-1: SVM

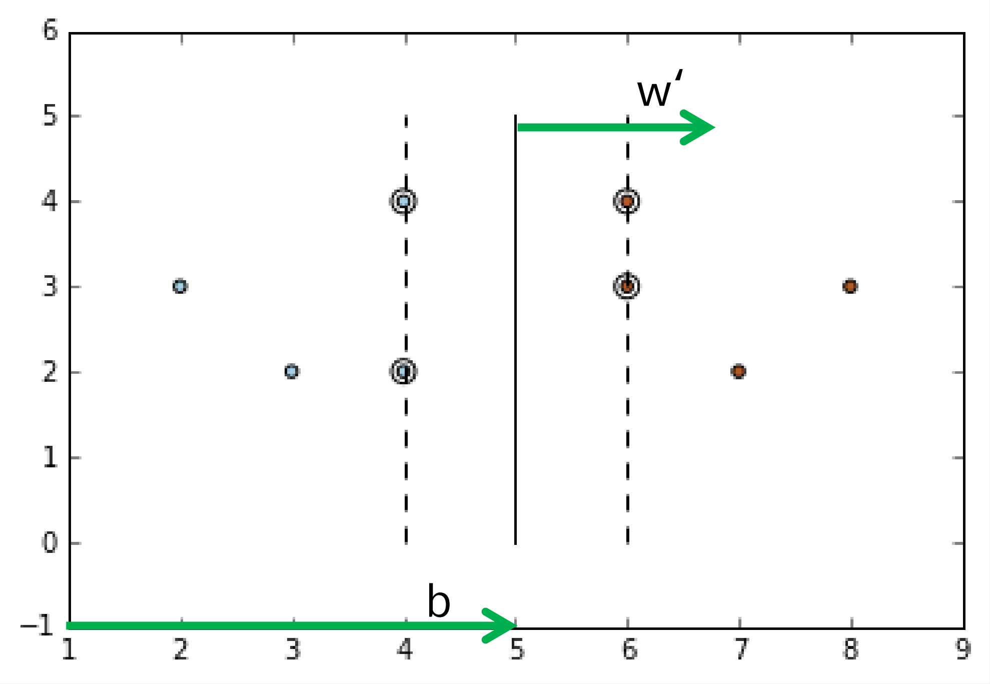
$x_1 = (2,3), x_2 = (3,2), x_3 = (4,4), x_4 = (4,2)$

$x_5 = (6,4), x_6 = (6,3), x_7 = (7,2), x_8 = (8,3)$

$A = \{x_1, x_2, x_3, x_4\}, B = \{x_5, x_6, x_7, x_8\}$

$$w' = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$b = -5$$



Exercise 10-1: SVM

$x_1 = (2,3), x_2 = (3,2), x_3 = (4,4), x_4 = (4,2)$

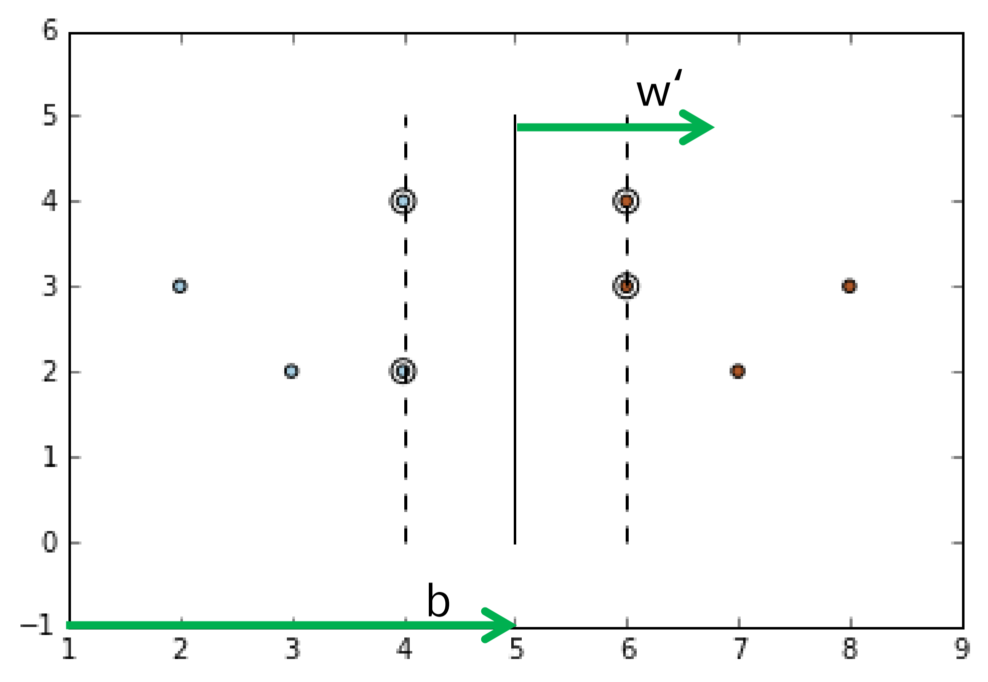
$x_5 = (6,4), x_6 = (6,3), x_7 = (7,2), x_8 = (8,3)$

$A = \{x_1, x_2, x_3, x_4\}, B = \{x_5, x_6, x_7, x_8\}$

$$w' = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$b = -5$$

$$h(x) = \langle w', x \rangle + b = x_1 - 5$$



Exercise 10-1: SVM

$x_1 = (2,3), x_2 = (3,2), x_3 = (4,4), x_4 = (4,2)$

$x_5 = (6,4), x_6 = (6,3), x_7 = (7,2), x_8 = (8,3)$

$A = \{x_1, x_2, x_3, x_4\}, B = \{x_5, x_6, x_7, x_8\}$

$$w' = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

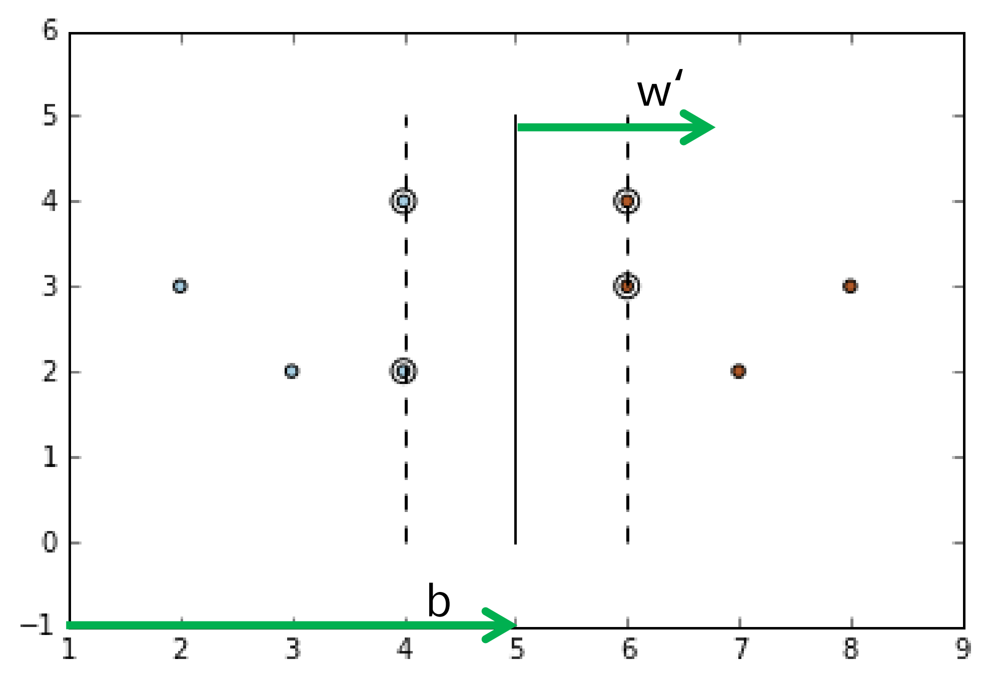
$$b = -5$$

$$h(x) = \langle w', x\rangle + b = x_1 - 5$$

$$h\left(\begin{pmatrix} 3 \\ 4 \end{pmatrix}\right) = 3 - 5 < 0$$

$$h\left(\begin{pmatrix} 7 \\ 4 \end{pmatrix}\right) = 7 - 5 > 0$$

$$h\left(\begin{pmatrix} 5 \\ 5 \end{pmatrix}\right) = 5 - 5 = 0$$



Exercise 10-2: Kernel Trick

To show: $K(x, y) = (x^T y + 1)^2 = \langle \phi(x), \phi(y) \rangle$

To show: $K(x,y) = (x^T y + 1)^2 = \langle \phi(x), \phi(y) \rangle$

$$\langle \phi(x), \phi(y) \rangle = \left(1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1 x_2\right)^{\mathrm{T}} \left(1, \sqrt{2}\, y_1, \sqrt{2}y_2, y_1^2, y_2^2, \sqrt{2}y_1 y_2\right)$$

$$K(x,y) = (x^T y + 1)^2 = (x_1 y_1 + x_2 y_2 + 1)^2$$

To show: $K(x, y) = (x^T y + 1)^2 = \langle \phi(x), \phi(y) \rangle$

$$\langle \phi(x), \phi(y) \rangle = \left(1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2\right)^{\mathrm{T}}\left(1, \sqrt{2}\, y_1, \sqrt{2}y_2, y_1^2, y_2^2, \sqrt{2}y_1y_2\right)$$
$$= 1 + 2x_1y_1 + 2x_2y_2 + x_1^2y_1^2 + x_2^2y_2^2 + 2x_1x_2y_1y_2$$

$$K(x, y) = (x^T y + 1)^2 = (x_1y_1 + x_2y_2 + 1)^2$$
$$= x_1y_1x_1y_1 + x_1y_1x_2y_2 + x_1y_1 + x_2y_2x_1y_1 + x_2y_2x_2y_2 + x_2y_2 + x_1y_1 + x_2y_2 + 1$$

To show: $K(x, y) = (x^T y + 1)^2 = \langle \phi(x), \phi(y) \rangle$

$$\langle \phi(x), \phi(y) \rangle = \left(1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2\right)^{\mathrm{T}} \left(1, \sqrt{2}\, y_1, \sqrt{2}y_2, y_1^2, y_2^2, \sqrt{2}y_1y_2\right)$$
$$= 1 + 2x_1y_1 + 2x_2y_2 + x_1^2y_1^2 + x_2^2y_2^2 + 2x_1x_2y_1y_2$$

$$K(x, y) = (x^T y + 1)^2 = (x_1y_1 + x_2y_2 + 1)^2$$
$$= x_1y_1x_1y_1 + x_1y_1x_2y_2 + x_1y_1 + x_2y_2x_1y_1 + x_2y_2x_2y_2 + x_2y_2 + x_1y_1 + x_2y_2 + 1$$

To show: $K(x, y) = (x^T y + 1)^2 = \langle \phi(x), \phi(y) \rangle$

$$\langle \phi(x), \phi(y) \rangle = \left(1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2\right)^{\mathrm{T}}\left(1, \sqrt{2}\, y_1, \sqrt{2}y_2, y_1^2, y_2^2, \sqrt{2}y_1y_2\right)$$
$$= 1 + 2x_1y_1 + 2x_2y_2 + x_1^2y_1^2 + x_2^2y_2^2 + 2x_1x_2y_1y_2$$

$$K(x, y) = (x^T y + 1)^2 = (x_1y_1 + x_2y_2 + 1)^2$$
$$= x_1y_1x_1y_1 + x_1y_1x_2y_2 + x_1y_1 + x_2y_2x_1y_1 + x_2y_2x_2y_2 + x_2y_2 + x_1y_1 + x_2y_2 + 1$$

Exercise 10-2: Kernel Trick

To show: $K(x,y) = (x^T y + 1)^2 = \langle \phi(x), \phi(y) \rangle$

$$\langle \phi(x), \phi(y) \rangle = \left(1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2\right)^{\mathrm{T}} \left(1, \sqrt{2}\, y_1, \sqrt{2}y_2, y_1^2, y_2^2, \sqrt{2}y_1y_2\right)$$
$$= 1 + 2x_1y_1 + 2x_2y_2 + x_1^2y_1^2 + x_2^2y_2^2 + 2x_1x_2y_1y_2$$

$$K(x,y) = (x^T y + 1)^2 = (x_1y_1 + x_2y_2 + 1)^2$$
$$= x_1y_1x_1y_1 + x_1y_1x_2y_2 + x_1y_1 + x_2y_2x_1y_1 + x_2y_2x_2y_2 + x_2y_2 + x_1y_1 + x_2y_2 + 1$$

Exercise 10-2: Kernel Trick

To show: $K(x,y) = (x^T y + 1)^2 = \langle \phi(x), \phi(y) \rangle$

$$\langle \phi(x), \phi(y) \rangle = \left(1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2\right)^{\mathrm{T}} \left(1, \sqrt{2}\,y_1, \sqrt{2}y_2, y_1^2, y_2^2, \sqrt{2}y_1y_2\right)$$
$$= 1 + 2x_1y_1 + 2x_2y_2 + x_1^2y_1^2 + x_2^2y_2^2 + 2x_1x_2y_1y_2$$

$$K(x,y) = (x^T y + 1)^2 = (x_1y_1 + x_2y_2 + 1)^2$$
$$= x_1y_1x_1y_1 + x_1y_1x_2y_2 + x_1y_1 + x_2y_2x_1y_1 + x_2y_2x_2y_2 + x_2y_2 + x_1y_1 + x_2y_2 + 1$$

To show: $K(x,y) = (x^T y + 1)^2 = \langle \phi(x), \phi(y) \rangle$

$$\langle \phi(x), \phi(y) \rangle = \left(1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2\right)^{\mathrm{T}} \left(1, \sqrt{2}\, y_1, \sqrt{2}y_2, y_1^2, y_2^2, \sqrt{2}y_1y_2\right)$$
$$= 1 + 2x_1y_1 + 2x_2y_2 + x_1^2y_1^2 + x_2^2y_2^2 + 2x_1x_2y_1y_2$$

$$K(x,y) = (x^T y + 1)^2 = (x_1y_1 + x_2y_2 + 1)^2$$
$$= x_1y_1x_1y_1 + x_1y_1x_2y_2 + x_1y_1 + x_2y_2x_1y_1 + x_2y_2x_2y_2 + x_2y_2 + x_1y_1 + x_2y_2 + 1$$

Exercise 10-2: Kernel Trick

To show: $K(x, y) = (x^T y + 1)^2 = \langle \phi(x), \phi(y) \rangle$

$$\langle \phi(x), \phi(y) \rangle = \left(1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2\right)^{\mathrm{T}}\left(1, \sqrt{2}\, y_1, \sqrt{2}y_2, y_1^2, y_2^2, \sqrt{2}y_1y_2\right)$$
$$= 1 + 2x_1y_1 + 2x_2y_2 + x_1^2y_1^2 + x_2^2y_2^2 + 2x_1x_2y_1y_2$$

$$K(x, y) = (x^T y + 1)^2 = (x_1y_1 + x_2y_2 + 1)^2$$
$$= x_1y_1x_1y_1 + x_1y_1x_2y_2 + x_1y_1 + x_2y_2x_1y_1 + x_2y_2x_2y_2 + x_2y_2 + x_1y_1 + x_2y_2 + 1$$

Gini-index and Information Gain

gift ID	gender of the recipient	useful	beautiful	self-made	eatable	liked by the recipient
1	male	yes	yes	no	yes	yes
2	male	yes	yes	yes	no	yes
3	male	yes	no	yes	no	yes
4	male	yes	yes	no	no	yes
5	male	no	yes	yes	yes	yes
6	male	no	little	no	yes	yes
7	male	no	no	yes	no	no
8	male	no	yes	no	no	no
9	female	no	yes	no	yes	yes
10	female	no	yes	no	no	yes
11	female	no	little	yes	no	yes
12	female	no	no	no	no	yes
13	female	no	little	yes	no	yes
14	female	no	little	yes	no	yes
15	female	no	little	no	yes	no
16	female	no	little	no	no	no

Exercise 10-3: Decision Tree

gift ID	gender of the recipient	useful	beautiful	self-made	eatable	liked by the recipient
1	male	yes	yes	no	yes	yes
2	male	yes	yes	yes	no	yes
3	male	yes	no	yes	no	yes
4	male	yes	yes	no	no	yes
5	male	no	yes	yes	yes	yes
6	male	no	little	no	yes	yes
7	male	no	no	yes	no	no
8	male	no	yes	no	no	no
9	female	no	yes	no	yes	yes
10	female	no	yes	no	no	yes
11	female	no	little	yes	no	yes
12	female	no	no	no	no	yes
13	female	no	little	yes	no	yes
14	female	no	little	yes	no	yes
15	female	no	little	no	yes	no
16	female	no	little	no	no	no

Exercise 10-3: Decision Tree

gift ID	gender of the recipient	useful	beautiful	self-made	eatable	liked by the recipient
1	male	yes	yes	no	yes	yes
2	male	yes	yes	yes	no	yes
3	male	yes	no	yes	no	yes
4	male	yes	yes	no	no	yes
5	male	no	yes	yes	yes	yes
6	male	no	little	no	yes	yes
7	male	no	no	yes	no	no
8	male	no	yes	no	no	no
9	female	no	yes	no	yes	yes
10	female	no	yes	no	no	yes
11	female	no	little	yes	no	yes
12	female	no	no	no	no	yes
13	female	no	little	yes	no	yes
14	female	no	little	yes	no	yes
15	female	no	little	no	yes	no
16	female	no	little	no	no	no
	Male: 6+,2- Female: 6+,2-	y: 4+,0- n: 8+,4-	y: 6+,1- n: 2+,1- l: 4+,2-	y: 6+,1- n: 6+,3-	y: 4+,1- n: 8+,3-	

Exercise 10-3: Decision Tree

Recursion 1	+	-	$gini(T_i)$	$gini_A(T)$
Gender: male Gender: female	6 6	2 2	0.375 0.375	0.3750
Useful: yes Useful: no	4 8	0 4	0.000 0.444	0.3333
Beautiful: yes Beauftiful: little Beautiful: no	6 4 2	1 2 1	0.245 0.444 0.444	0.3571
Self-made: yes Self-made: no	6 6	1 3	0.245 0.444	0.3571
Eatable: yes Eatable: no	4 8	1 3	0.320 0.397	0.3727

$$gini(T) = 1 - \sum_{j=1}^{k} p_j^2 \qquad gini_A(T) = \sum_{i=1}^{m} \frac{|T_i|}{|T|} \cdot gini(T_i)$$

Exercise 10-3: Decision Tree

Recursion 2	+	-	$gini(T_i)$	$gini_A(T)$
Gender: male Gender: female	2 6	2 2	0.500 0.375	0.4167
Beautiful: yes Beauftiful: little Beautiful: no	3 4 1	1 2 1	0.375 0.444 0.500	0.4306
Self-made: yes Self-made: no	4 4	1 3	0.320 0.490	0.4190
Eatable: yes Eatable: no	3 5	1 3	0.375 0.469	0.4375

$$gini(T) = 1 - \sum_{j=1}^{k} p_j^2$$

$$gini_A(T) = \sum_{i=1}^{m} \frac{|T_i|}{|T|} \cdot gini(T_i)$$

Recursion 3 (male)	+	-	$gini(T_i)$	$gini_A(T)$
Beautiful: yes Beauftiful: little Beautiful: no	1 1 0	1 0 1	0.500 0.000 0.000	0.2500
Self-made: yes Self-made: no	1 1	1 1	0.500 0.500	0.5000
Eatable: yes Eatable: no	2 0	0 2	0.000 0.000	0.0000

$$gini(T) = 1 - \sum_{j=1}^{k} p_j^2$$

$$gini_A(T) = \sum_{i=1}^{m} \frac{|T_i|}{|T|} \cdot gini(T_i)$$

Exercise 10-3: Decision Tree

Recursion 3 (female)	+	-	$gini(T_i)$	$gini_A(T)$
Beautiful: yes Beauftiful: little Beautiful: no	2 3 1	0 2 0	0.000 0.480 0.000	0.3000
Self-made: yes Self-made: no	3 3	0 2	0.000 0.480	0.3000
Eatable: yes Eatable: no	1 5	1 1	0.500 0.278	0.3333

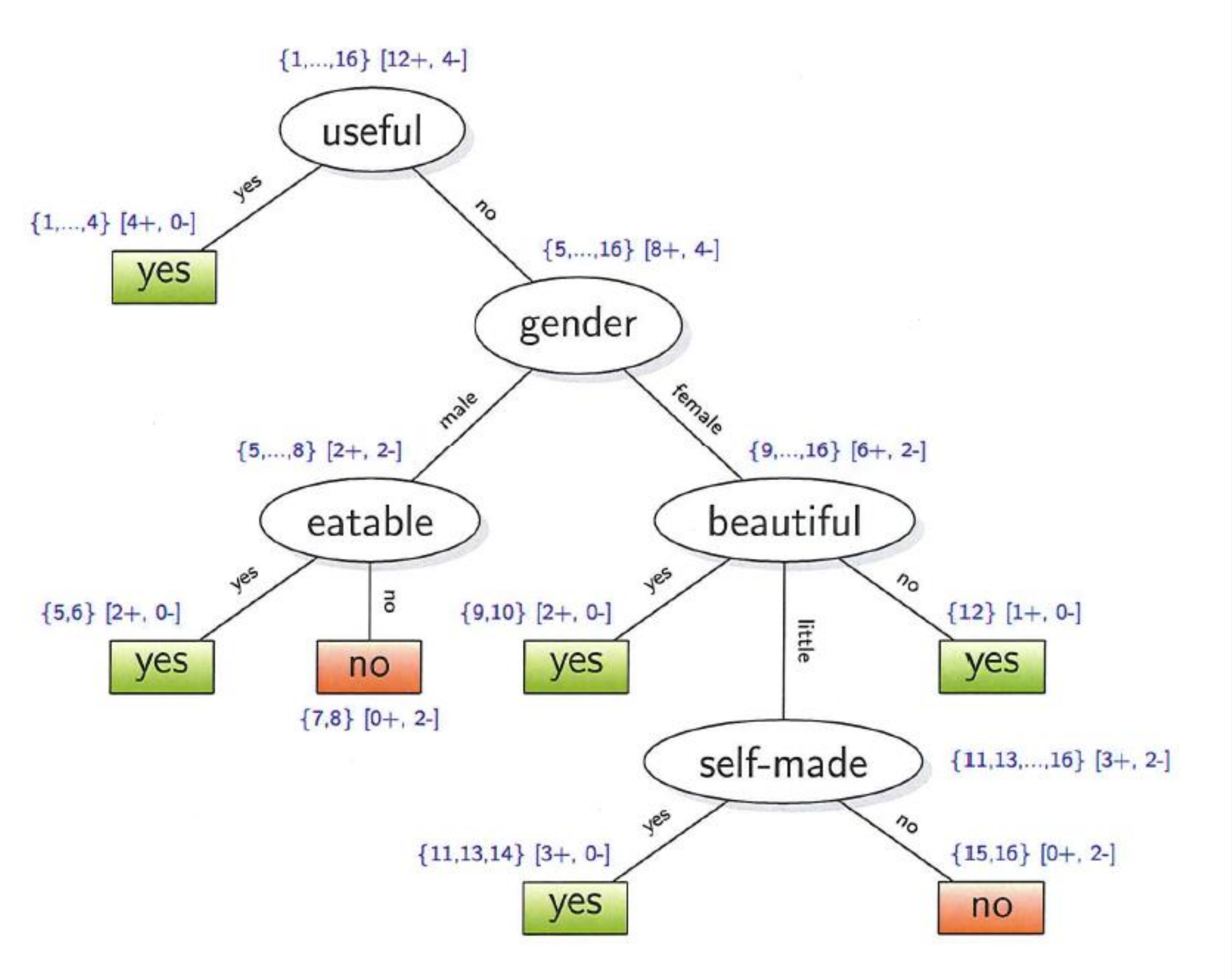
$$gini(T) = 1 - \sum_{j=1}^{k} p_j^2$$

$$gini_A(T) = \sum_{i=1}^{m} \frac{|T_i|}{|T|} \cdot gini(T_i)$$

Recursion 4	+	-	$gini(T_i)$	$gini_A(T)$
Self-made: yes Self-made: no	3 0	0 2	0.000 0.000	0.0000
Eatable: yes Eatable: no	0 3	1 1	0.000 0.375	0.3000

$$gini(T) = 1 - \sum_{j=1}^{k} p_j^2 \qquad gini_A(T) = \sum_{i=1}^{m} \frac{|T_i|}{|T|} \cdot gini(T_i)$$

Exercise 10-3: Decision Tree



Exercise 10-3: Decision Tree

Recursion 1	+	-	$entropy(T_i)$	$IG(T)$
Gender: male Gender: female	6 6	2 2	0.811 0.811	0.0000
Useful: yes Useful: no	4 8	0 4	0.000 0.918	0.1225
Beautiful: yes Beauftiful: little Beautiful: no	6 4 2	1 2 1	0.592 0.918 0.918	0.0356
Self-made: yes Self-made: no	6 6	1 3	0.592 0.918	0.0356
Eatable: yes Eatable: no	4 8	1 3	0.722 0.845	0.0044

$$entropy(T) = -\sum_{i=1}^{k} p_i \cdot \log_2 p_i$$

$$IG(T,A) = entropy(T) - \sum_{i=1}^{m} \frac{|T_i|}{|T|} \cdot entropy(T_i)$$

Exercise 10-3: Decision Tree

Recursion 2	+	-	$entropy(T_i)$	$IG(T)$
Gender: male Gender: female	2 6	2 2	1.000 0.811	0.0440
Beautiful: yes Beauftiful: little Beautiful: no	3 4 1	1 2 1	0.811 0.918 1.000	0.0220
Self-made: yes Self-made: no	4 4	1 3	0.722 0.985	0.0426
Eatable: yes Eatable: no	3 5	1 3	0.811 0.954	0.0117

$$entropy(T) = -\sum_{i=1}^{k} p_i \cdot \log_2 p_i$$

$$IG(T, A) = entropy(T) - \sum_{i=1}^{m} \frac{|T_i|}{|T|} \cdot entropy(T_i)$$

Exercise 10-3: Decision Tree

Recursion 3a (male)	+	-	$entropy(T_i)$	$IG(T)$
Beautiful: yes Beauftiful: little Beautiful: no	1 1 0	1 0 1	1.000 0.000 0.000	0.5000
Self-made: yes Self-made: no	1 1	1 1	1.000 1.000	0.0000
Eatable: yes Eatable: no	2 0	0 2	0.000 0.000	1.0000

$$entropy(T) = -\sum_{i=1}^{k} p_i \cdot \log_2 p_i$$

$$IG(T,A) = entropy(T) - \sum_{i=1}^{m} \frac{|T_i|}{|T|} \cdot entropy(T_i)$$

Recursion 3b (female)	+	-	$entropy(T_i)$	$IG(T)$
Beautiful: yes Beauftiful: little Beautiful: no	2 3 1	0 2 0	0.000 0.971 0.000	0.2041
Self-made: yes Self-made: no	3 3	0 2	0.000 0.971	0.2041
Eatable: yes Eatable: no	1 5	1 1	1.000 0.650	0.0735

$$entropy(T) = -\sum_{i=1}^{k} p_i \cdot \log_2 p_i$$

$$IG(T, A) = entropy(T) - \sum_{i=1}^{m} \frac{|T_i|}{|T|} \cdot entropy(T_i)$$

Exercise 10-3: Decision Tree

Recursion 4	+	-	$entropy(T_i)$	$IG(T)$
Self-made: yes Self-made: no	3 0	0 2	0.000 0.000	0.9710
Eatable: yes Eatable: no	0 3	1 1	0.000 0.811	0.3222

$$entropy(T) = -\sum_{i=1}^{k} p_i \cdot \log_2 p_i$$

$$IG(T, A) = entropy(T) - \sum_{i=1}^{m} \frac{|T_i|}{|T|} \cdot entropy(T_i)$$

Exercise 10-3: Decision Tree