Aprendizagem 2021/22

Homework II – Group 17

I. Pen-and-paper

1) I

1) $E(w) = \frac{1}{2}\sum_{i=1}^{n}(z_i - f(x_i, w))^2 = \frac{1}{2}\|z - \hat{z}\|^2 = \frac{1}{2}(z - Xw)^T(z - Xw)$

$$\nabla = \left[\frac{\partial}{\partial w_1} \quad \frac{\partial}{\partial w_2} \quad \cdots \quad \frac{\partial}{\partial w_m}\right]^T$$

$$\nabla E(w) = \left[\frac{\partial E}{\partial w_1} \quad \frac{\partial E}{\partial w_2} \quad \cdots \quad \frac{\partial E}{\partial w_m}\right]^T = \nabla\left(\frac{1}{2}(z - X.w)^T(z - X.w)\right) = 0$$

$$\Leftrightarrow \frac{1}{2}\nabla\left((z - Xw)^T(z - Xw)\right) = 0$$

$$\Leftrightarrow \nabla(z^t.z - 2z^T.X.w + w^T.X^T.X.w) = 0$$

$$\Leftrightarrow \nabla(z^t.z) - 2\nabla(z^T.X.w) + \nabla(w^T.X^T.X.w) = 0$$

$$\Leftrightarrow -2X^Tz + 2X^T.X.w = 0$$

$$\Leftrightarrow X^T.z = X^T.X.w$$

$$\Leftrightarrow (X^t.X)^{-1}.X^T.z = w$$

$$\Leftrightarrow w = (X^T.X)^{-1}.X^T.z$$

$$f(x, w) = \sum_{j=0}^{3} w_j.\phi_j(x)$$

$$= w_0.\phi_0(x) + w_1\phi_1(x) + w_2\phi_2(x) + w_3\phi_3(x) \quad , \quad \phi_j(x) = \|X\|_2^j$$

$$X = \begin{array}{c} \begin{array}{cccc} y_0 & y_1 & y_2 & y_3 \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 5 \\ 1 & 0 & 2 & 4 \\ 1 & 1 & 2 & 3 \\ 1 & 2 & 0 & 7 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 2 \\ 1 & 0 & 2 & 9 \end{pmatrix} \end{array} ; \quad \phi(X) = \begin{array}{c} \begin{array}{cccc} y_0 & y_1 & y_2 & y_3 \end{array} \\ \begin{pmatrix} 1 & \sqrt{2} & 2 & \sqrt{8} \\ 1 & \sqrt{27} & 27 & \sqrt{19683} \\ 1 & \sqrt{20} & 20 & \sqrt{8000} \\ 1 & \sqrt{14} & 14 & \sqrt{2744} \\ 1 & \sqrt{53} & 53 & \sqrt{148877} \\ 1 & \sqrt{3} & 3 & \sqrt{27} \\ 1 & \sqrt{8} & 8 & \sqrt{512} \\ 1 & \sqrt{85} & 85 & \sqrt{614125} \end{pmatrix} \end{array} ; \quad z = \begin{pmatrix} 1 \\ 3 \\ 2 \\ 0 \\ 6 \\ 4 \\ 5 \\ 7 \end{pmatrix}$$

$\|x_1\| = \sqrt{2}$ $\qquad$ $\|x_6\| = \sqrt{3}$

$\|x_2\| = \sqrt{27}$ $\qquad$ $\|x_7\| = \sqrt{8}$

$\|x_3\| = \sqrt{20}$ $\qquad$ $\|x_8\| = \sqrt{85}$

$\|x_4\| = \sqrt{14}$

$\|x_5\| = \sqrt{53}$

$$w = (\phi(x)^T.\phi(x))^{-1}.\phi(x)^T.z$$

$$w = \left(\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \sqrt{2} & \sqrt{27} & \sqrt{20} & \sqrt{14} & \sqrt{53} & \sqrt{3} & \sqrt{8} & \sqrt{85} \\ 2 & 27 & 20 & 14 & 53 & 3 & 8 & 85 \\ \sqrt{8} & \sqrt{19683} & \sqrt{8000} & \sqrt{2744} & \sqrt{148877} & \sqrt{27} & \sqrt{512} & \sqrt{614125} \end{bmatrix} . \begin{bmatrix} 1 & \sqrt{2} & 2 & \sqrt{8} \\ 1 & \sqrt{27} & 27 & \sqrt{19683} \\ 1 & \sqrt{20} & 20 & \sqrt{8000} \\ 1 & \sqrt{14} & 14 & \sqrt{2744} \\ 1 & \sqrt{53} & 53 & \sqrt{148877} \\ 1 & \sqrt{3} & 3 & \sqrt{27} \\ 1 & \sqrt{8} & 8 & \sqrt{512} \\ 1 & \sqrt{85} & 85 & \sqrt{614125} \end{bmatrix}\right)^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \sqrt{2} & \sqrt{27} & \sqrt{20} & \sqrt{14} & \sqrt{53} & \sqrt{3} & \sqrt{8} & \sqrt{85} \\ 2 & 27 & 20 & 14 & 53 & 3 & 8 & 85 \\ \sqrt{8} & \sqrt{19683} & \sqrt{8000} & \sqrt{2744} & \sqrt{148877} & \sqrt{27} & \sqrt{512} & \sqrt{614125} \end{bmatrix} . \begin{bmatrix} 1 \\ 3 \\ 2 \\ 0 \\ 6 \\ 4 \\ 5 \\ 7 \end{bmatrix}$$

$$w = \begin{bmatrix} 8,20 & -6,23 & 1,30 & -7,83 \times 10^{-2} \\ -6,23 & 5,08 & -1,10 & 6,86 \times 10^{-2} \\ 1,30 & -1,10 & 2,47 \times 10^{-1} & -1,57 \times 10^{-2} \\ -7,93 \times 10^{-2} & 6,86 \times 10^{-2} & -1,57 \times 10^{-2} & 1,01 \times 10^{-3} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \sqrt{2} & \sqrt{27} & \sqrt{20} & \sqrt{14} & \sqrt{53} & \sqrt{3} & \sqrt{8} & \sqrt{85} \\ 2 & 27 & 20 & 14 & 53 & 3 & 8 & 85 \\ \sqrt{8} & \sqrt{19683} & \sqrt{8000} & \sqrt{2744} & \sqrt{148877} & \sqrt{27} & \sqrt{512} & \sqrt{614125} \end{bmatrix} . \begin{bmatrix} 1 \\ 3 \\ 2 \\ 0 \\ 6 \\ 4 \\ 5 \\ 7 \end{bmatrix}$$

Aprendizagem 2021/22

Homework II – Group 17

$$w = \begin{bmatrix} 1,77 & -0,08 & -0,67 & -1,01 & 1,38 & 0,91 & -0,79 & -0,51 \\ -1,06 & -0,032 & 0,53 & 0,90 & -1,31 & -0,39 & 0,85 & 0,51 \\ 0,19 & 0,04 & -0,09 & -0,19 & 0,32 & 0,05 & -0,20 & -0,14 \\ -0,011 & -0,0043 & 0,0044 & 0,011 & -0,02 & -0,0022 & 0,012 & 0,011 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 3 \\ 2 \\ 0 \\ 6 \\ 4 \\ 5 \\ 7 \end{bmatrix}$$

$$w = \begin{bmatrix} 4,584 \\ -1,687 \\ 0,338 \\ -0,013 \end{bmatrix}$$

2) 2) $\phi(x_9) = \begin{bmatrix} 1 \\ 2 \\ 4 \\ 8 \end{bmatrix}$; $\phi(x_{10}) = \begin{bmatrix} 1 \\ \sqrt{6} \\ 6 \\ \sqrt{216} \end{bmatrix}$

$$f(x_{10}, w) = w^T \cdot \phi(x_{10}) = [4,584 \quad -1,687 \quad 0,338 \quad -0,013] \begin{bmatrix} 1 \\ \sqrt{6} \\ 6 \\ \sqrt{216} \end{bmatrix} = 2,2816$$

$$f(x_9, w) = w^T \cdot \phi(x_9) = [4,584 \quad -1,687 \quad 0,338 \quad -0,013] \begin{bmatrix} 1 \\ 2 \\ 4 \\ 8 \end{bmatrix} = 2,4536$$

$$RMSE = \sqrt{\frac{\sum (z - \hat{z})^2}{n}} = \sqrt{\frac{(2 - 2,4536)^2 + (4 - 2,2816)^2}{2}} = 1,2567$$

3) 3)

	y_1	y_2	y_3	z
x_1	1	1	0	N
x_2	1	1	1	N
x_3	0	2	1	N
x_4	1	2	1	N
x_5	2	0	1	P
x_6	1	1	0	P
x_7	2	0	0	P
x_8	0	2	1	P
x_9	2	0	0	N
x_{10}	1	2	0	P

mediana $= \frac{2+3}{2} = 2,5$

$IG(z \mid y_1) = H(z) - H(z \mid y_1)$

$IG(z \mid y_2) = H(z) - H(z \mid y_2)$

$IG(z \mid y_3) = H(z) - H(z \mid y_3)$

$$H(z) = -\sum_z P(z=z) \cdot \log_2 P(z=z) = -\left(2 \cdot \frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right)\right) = \log_2(2) = 1$$

$$P(z) = \begin{cases} \frac{1}{2}, & z = N \\ \frac{1}{2}, & z = P \end{cases}$$

$$P(y_1) = \begin{cases} \frac{1}{4}, & y_1 = 0 \\ \frac{1}{2}, & y_1 = 1 \\ 1/4, & y_1 = 2 \end{cases} \qquad P(y_2) = \begin{cases} 1/4, & y_1 = 0 \\ 3/8, & y_2 = 1 \\ 3/8, & y_2 = 2 \end{cases} \qquad P(y_3) = \begin{cases} \frac{3}{8}, & y_3 = 0 \\ \frac{5}{8}, & y_3 = 1 \end{cases}$$

$$H(z \mid y_1 = 0) = -\left(P(z=N \mid y_1=0) \cdot \log_2(P(z=N \mid y_1=0)) + P(z=P \mid y_1=0) \cdot \log_2(P(z=P \mid y_1=0))\right) = -\left(1/2 \cdot \log(1/2) \times 2\right) = \log_2 2 = 1$$

$$H(z \mid y_1 = 1) = -\left(P(z=N \mid y_1=1) \cdot \log_2(P(z=N \mid y_1=1)) + P(z=P \mid y_1=1) \cdot \log_2(P(z=P \mid y_1=1))\right) = -\left(3/4 \cdot \log_2(3/4) + 1/4 \log_2(1/4)\right) = 0,8113$$

$$H(z \mid y_1 = 2) = -\left(P(z=N \mid y_1=2) \cdot \log_2(P(z=N \mid y_1=2)) + P(z=P \mid y_1=2) \cdot \log_2(P(z=P \mid y_1=2))\right) = -(0 + 1 \cdot \log_2(1)) = 0$$

$$\begin{aligned} H(z \mid y_1) &= \frac{1}{4} H(z \mid y_1 = 0) + \frac{1}{2} H(z \mid y_1 = 1) + \frac{1}{4} H(z \mid y_1 = 2) \\ &= \frac{1}{4} + \frac{1}{2} \cdot 0,8113 + 0 \\ &= 0,65565 \end{aligned}$$

$IG(z \mid y_1) = 1 - 0,65565 = 0,34435$

$H(z \mid y_2 = 0) = (0 + 1 \cdot \log_2 1) = 0$

$H(z \mid y_2 = 1) = -\left(\frac{2}{3} \cdot \log_2 \frac{2}{3} + \frac{1}{3} \cdot \log_2 \frac{1}{3}\right) = 0,9183$

$H(z \mid y_2 = 2) = -\left(\frac{2}{3} \cdot \log_2 \frac{2}{3} + \frac{1}{3} \log_2 \frac{1}{3}\right) = 0,9183$

$H(z \mid y_2) = \frac{1}{4} \cdot 0 + \frac{3}{8}(0,9183) \times 2 = \frac{3}{4} \cdot 0,9183 = 0,688725$

$I(z \mid y_2) = 1 - 0,688725 = 0,311275$

$H(z|y_3=0) = -\left(\frac{1}{3}\log_2\left(\frac{1}{3}\right) + \frac{2}{3}\log_2\left(\frac{2}{3}\right)\right) = 0{,}9183$

$H(z|y_3=1) = -\left(\frac{3}{5}\log_2\left(\frac{3}{5}\right) + \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right) = 0{,}9710$

$H(z|y_3) = \frac{3}{8}\cdot 0{,}9183 + \frac{5}{8}\cdot 0{,}9710 = 0{,}9512$

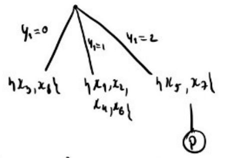
$IG(z|y_3) = 1 - 0{,}9512 = 0{,}0488$

Logo como $IG(z|y_1) > IG(z|y_2) > IG(z|y_3)$ escolhe-se y_1 para a raiz da árvore.

$X_0 = \{x_3, x_8\}$

$X_1 = \{x_1, x_2, x_4, x_6\}$

$X_2 = \{x_5, x_7\}$



$IG(z|y_2, y_1=0) = H(z|y_1=0) - H(z|y_2, y_1=0) = 1 - 1 = 0$

$IG(z|y_3, y_1=0) = H(z|y_1=0) - H(z|y_3, y_1=0) = 1 - 1 = 0$

$H(z|y_2, y_1=0) = H(z|y_2=2, y_1=0) = -\big(P(z=N|y_2=2,y_1=0)\cdot\log_2(P(z=N|y_2=2,y_1=0)) + P(z=P|y_2=2,y_1=0)\cdot\log_2(P(z=P|y_2=2,y_1=0))\big) = 1$

$H(z|y_3=0, y_1=0) = 0$

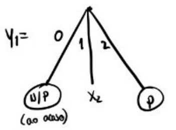
$H(z|y_3=1, y_1=0) = 1$

$H(z|y_3, y_1=0) = 1$

$IG(z|y_3, y_1) = 0$

$P(y_2|y_1=0) = \{1, y_2=2\}$

$P(y_3|y_1=0) = \{1, y_3=1\}$



, $X_2 = \{x_1, x_2, x_4, x_6\}$

$IG(z|y_2, y_1=1) = H(z|y_1=1) - H(z|y_2, y_1=1) = 0{,}8113 - 0{,}688725 = 0{,}122575$

$IG(z|y_3, y_1=1) = H(z|y_1=1) - H(z|y_3, y_1=1) = 0{,}8113 - 0{,}5 = 0{,}3113$

> logo, escolhemos y_3.

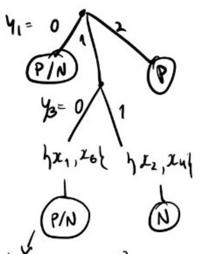
$P(y_2|y_1=1) = \begin{cases} \frac{3}{4}, & y_2=1 \\ 1/4, & y_2=2 \end{cases}$

$P(y_3|y_1=1) = \begin{cases} \frac{1}{2}, & y_3=0 \\ \frac{1}{2}, & y_3=1 \end{cases}$

$H(z|y_3, y_1=1) = \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 0 = \frac{1}{2}$

$H(z|y_3=0, y_1=1) = 1$

$H(z|y_3=1, y_1=1) = 0$



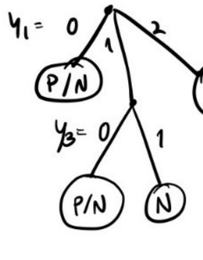
(Por observação)

$H(z|y_2=1, y_1=1) = -\left(\frac{2}{3}\log_2\left(\frac{2}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right) = 0{,}9183$

$H(z|y_2=2, y_1=1) = -\left(1\cdot\log_2(1)\right) = 0$

$H(z|y_2, y_1=1) = \frac{3}{4}\cdot 0{,}9183 = 0{,}688725$

4) 4)



Ao percorrer a árvore com x_9 prevê-se P.

Ao percorrer a árvore com x_{10} não há uma classificação bem definida

Accuracy: No melhor caso: 50%.

No pior caso: 0%.

Ponderado: 25%.

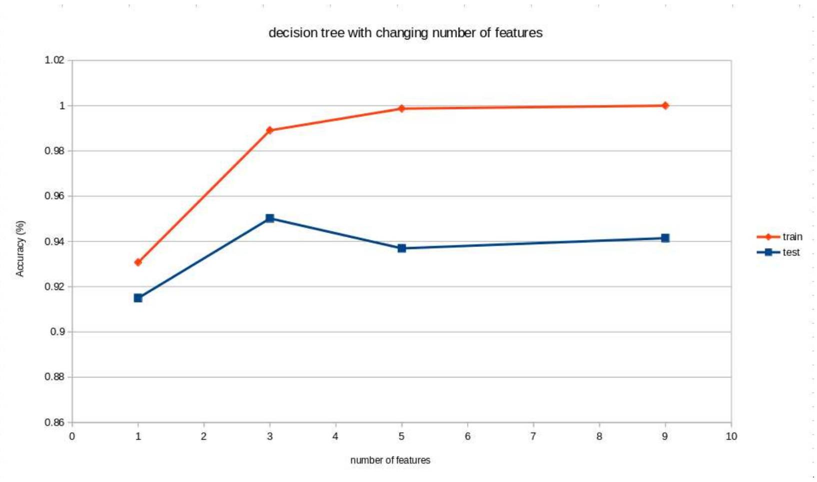
Aprendizagem 2021/22

Homework II – Group 17

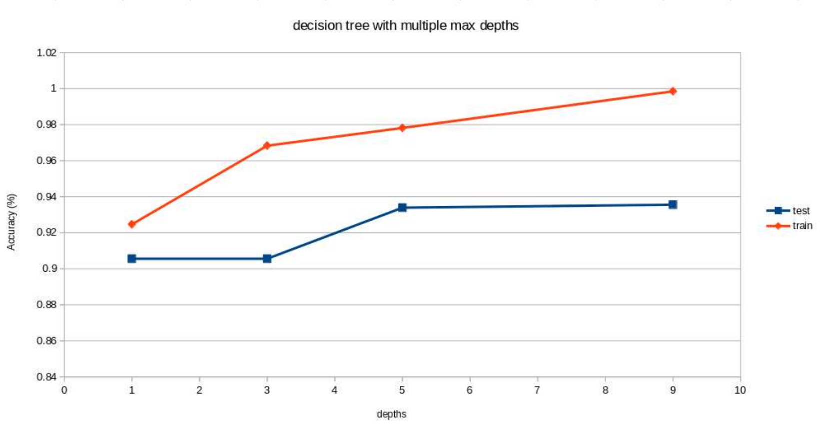
II. Programming and critical analysis

5)

i.



ii.



6) Uma vez que as decision trees são um algoritmo de local learning, estão mais sujeitas a overfitting, tendo assim em ambos os gráficos uma melhor performance no treino do que no teste; para além disso, verifica-se que, de modo geral, a variação da accuracy é bastante semelhante em ambos os gráficos, concluindo-se assim que o número de features selecionadas e a profundidade máxima da árvore influenciam do mesmo modo a performance do classificador.

7) Depth = 5, uma vez que obtém resultados semelhantes a depth = 9 e, para além disso, por ser um valor mais baixo está menos sujeito a overfitting.

III. APPENDIX

```
import scipy.io.arff
import numpy as np
from sklearn.model_selection import KFold
from sklearn.feature_selection import SelectKBest, mutual_info_classif
from sklearn.tree import DecisionTreeClassifier
from copy import deepcopy

FILE_NAME = "../hw1/breast.w.new.arff" # ficheiro sem observacoes com entradas "?"
arff = scipy.io.arff
SEED = 17
data, meta = arff.loadarff(FILE_NAME)
k_fold = KFold(n_splits=10, shuffle=True, random_state=SEED)
max_features = { 1 : None, 3 : None, 5 : None, 9 : None }
max_features_folds = list()
test_max_f_accs = list(0 for i in range(4))
train_max_f_accs = list(0 for i in range(4))
test_max_d_accs = list(0 for i in range(4))
train_max_d_accs = list(0 for i in range(4))
max_depth_folds = list()
max_depth = deepcopy(max_features)
kbests = list()
targets = list(el["Class"] for el in data)
training_data = np.array(list(list(el[i] for i in range(9)) for el in data))

for n in (1,3,5,9):
    kbests.append(SelectKBest(mutual_info_classif, k=n).fit_transform(training_data, targets))
    max_features[n] = DecisionTreeClassifier()
    max_depth[n] = DecisionTreeClassifier(max_depth=n)

i = 0
for n in (1, 3, 5, 9):
    for train, test in k_fold.split(kbests[i]):
        max_features[n].fit(np.array([kbests[i][j] for j in train]),\
                            np.array([targets[j] for j in train]))
        test_acc = max_features[n].score(np.array([kbests[i][j] for j in test]),\
                            np.array([targets[j] for j in test]), sample_weight=None)
        train_acc = max_features[n].score(np.array([kbests[i][j] for j in train]),\
                            np.array([targets[j] for j in train]))
        test_max_f_accs[i] += test_acc
        train_max_f_accs[i] += train_acc

    i += 1

i = 0
for n in (1, 3, 5, 9):
    for train, test in k_fold.split(train):
        max_depth[n].fit(np.array([training_data[i] for i in train]),\
                            np.array([targets[i] for i in train]))
        test_acc = max_depth[n].score(np.array([training_data[i] for i in test]),\
                            np.array([targets[i] for i in test]), sample_weight=None)
        train_acc = max_depth[n].score(np.array([training_data[i] for i in train]),\
                            np.array([targets[i] for i in train]))
        test_max_d_accs[i] += test_acc
        train_max_d_accs[i] += train_acc

    i += 1

test_max_f_accs = list(el / 10 for el in test_max_f_accs)
train_max_f_accs = list(el / 10 for el in train_max_f_accs)
test_max_d_accs = list(el / 10 for el in test_max_d_accs)
train_max_d_accs = list(el / 10 for el in train_max_d_accs)
```

END