שאלה 1:

א) WL הינה משפחה של לומדים חלשים כפי שלמדנו בהרצאה.
אנו מניחים שעבור כל התפלגות $D^{(t)}$ על סט האימון S
ניתן למצוא לומד חלש עבורו מתקיים:

$$\epsilon_t = \sum_{i=1}^{m} D_i^{(t)} \cdot 1[y_i \neq h_t(x_i)] \le \frac{1}{2}$$

כלומר, שגיאת האימון המשוקללת עבור כל איטרציה t יותר טובה מניחוש אקראי.

לפי צעדי האלגוריתם נאתחל $D^{(1)} = (\frac{1}{m}, \dots, \frac{1}{m})$ - חלוקה יוניפורמית. בכל צעד t נבצע:

1) נמצא לומד חלש $h_t \in WL(D^{(t)}, S)$ המקיים את ההנחה.

2) נחשב את השגיאה המשוקללת ϵ_t לפי $D^{(t)}$ על S.

3) נחשב את המשקל של הלומד החלש h_t ע"י:
$w_t = \frac{1}{2}\log(\frac{1}{\epsilon_t} - 1)$ המשקל ביחס הפוך ל- ϵ_t.

4) נעדכן המשקלים עבור האיטרציה הבאה ע"י:

$$D_i^{(t+1)} = \frac{D_i^{(t)} \cdot \exp(-w_t \cdot y_i \cdot h_t(x_i))}{\sum_{j=1}^{m} D_j^{(t)} \cdot \exp(-w_t \cdot y_j \cdot h_t(x_j))}$$

עבור פוינט שטעינו בסיווגה, נגדיל את משקלה כדי לנסות לספק תוצאה טובה יותר באיטרציה הבאה.

5) נסכום את כל המסווגים המשוקללים ונמצא את הסיווג.

ב)

$$\sum_{i=1}^{m} D_i^{(t)} \cdot e^{-y_i w_t h_t(x_i)} =$$

$$= \sum_{\substack{i=1 \\ y_i = h_t(x_i)}}^{m} D_i^{(t)} \cdot e^{-y_i w_t h_t(x_i)} + \sum_{\substack{i=1 \\ y_i \neq h_t(x_i)}}^{m} D_i^{(t)} \cdot e^{-y_i w_t h_t(x_i)} =$$

$$= \sum_{\substack{i=1 \\ y_i = h_t(x_i)}}^{m} D_i^{(t)} \cdot e^{-w_t} + \sum_{\substack{i=1 \\ y_i \neq h_t(x_i)}}^{m} D_i^{(t)} \cdot e^{w_t} =$$

$$= e^{-w_t}\cdot \sum_{\substack{i=1\\ y_i=h_t(x_i)}}^{m} D_i^{(t)} + e^{w_t}\cdot \sum_{\substack{i=1\\ y_i\neq h_t(x_i)}}^{m} D_i^{(t)} =$$

$$\underset{\text{מהגדרת } \varepsilon_t}{=} e^{-w_t}\cdot(1-\varepsilon_t) + e^{w_t}\cdot\varepsilon_t = e^{w_t}\cdot\varepsilon_t + \frac{(1-\varepsilon_t)}{e^{w_t}} =$$

$$\underset{\text{מהגדרת } w_t}{=} \exp\left(\tfrac{1}{2}\log\left(\tfrac{1}{\varepsilon_t}-1\right)\right)\cdot\varepsilon_t + \frac{(1-\varepsilon_t)}{\exp\left(\tfrac{1}{2}\log\left(\tfrac{1}{\varepsilon_t}-1\right)\right)} =$$

$$= \exp\left(\log\left(\sqrt{\tfrac{1-\varepsilon_t}{\varepsilon_t}}\right)\right)\cdot\varepsilon_t + (1-\varepsilon_t)\cdot\frac{1}{\exp\left(\log\left(\sqrt{\tfrac{1-\varepsilon_t}{\varepsilon_t}}\right)\right)} =$$

$$= \sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}}\cdot\varepsilon_t + \sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}}\cdot(1-\varepsilon_t) = 2\sqrt{\varepsilon_t(1-\varepsilon_t)}$$

$\blacksquare$

סע' 1) נחלק למקרים:

I $y_i=\hat{h}(x_i) \Rightarrow 1[y_i\neq\hat{h}(x_i)]=0\le e^{-y_i f_T(x_i)}$

אם החיזוי נכון אז האינדיקטור יהיה 0 כיוון שבפקט

II $y_i\neq\hat{h}(x_i) \Rightarrow 1[y_i\neq\hat{h}(x_i)]=1$

מכיוון $\hat{h}(x_i)=\mathrm{sign}(f_T(x_i)) \Rightarrow y_i f_T(x_i)\le 0$

$\Rightarrow e^{-y_i\cdot f_T(x_i)}\ge e^0=1=1[y_i\neq\hat{h}(x_i)]$

$\Rightarrow \forall i\in[m]: 1[y_i\neq\hat{h}(x_i)]\le e^{-y_i\cdot f_T(x_i)}$

$\Rightarrow L_S(\hat{h})=\frac{1}{m}\sum_{i=1}^{m}1[y_i\neq\hat{h}(x_i)]\le\frac{1}{m}\sum_{i=1}^{m}e^{-y_i f_T(x_i)}=Z_T$

2) בסיס האינדוקציה: עבור $t=1$, כיוון ש- $f_0(x_i)\equiv 0$

מאותחלות לאחידות מתקיים:

$$\forall i: D_i^{(1)}=\frac{1}{m}=\frac{e^{-y_i\cdot 0}}{\sum_{j=1}^{m}e^{-y_j\cdot 0}}=\frac{e^{-y_i\cdot f_0(x_i)}}{\sum_{j=1}^{m}e^{-y_j\cdot f_0(x_j)}}\longrightarrow D_i^{(1)}\propto e^{-y_i f_0(x_i)}$$

צעד האינדוקציה: נניח נכונות ל- $(t-1)$ ונוכיח t:

מצעד העדכון של האלג' נקבל ש:

$$D_i^{(t+1)}\propto D_i^{(t)}\cdot e^{-y_i\cdot w_t\cdot h_t(x_i)}\propto e^{-y_i\cdot f_{t-1}(x_i)}\cdot e^{-y_i\cdot w_t\cdot h_t(x_i)}=$$

$$= e^{-y_i(f_{t-1}(x_i)+w_t h_t(x_i))} \underset{\uparrow}{=} e^{-y_i \cdot f_t(x_i)}$$

הנחת האינדוקציה

(3

מההגדרה / מההגדרה

$$\frac{Z_t}{Z_{t-1}} \overset{\downarrow}{=} \frac{1}{m}\sum_{i=1}^{m} e^{-y_i f_t(x_i)} \cdot \frac{1}{\frac{1}{m}\sum_{i=1}^{m} e^{-y_i \cdot f_{t-1}(x_i)}} =$$

$$= \frac{\sum_{i=1}^{m} e^{-y_i\cdot(f_{t-1}(x_i)+w_t h_t(x_i))}}{\sum_{i=1}^{m} e^{-y_i \cdot f_{t-1}(x_i)}} =$$

$$= \frac{\sum_{i=1}^{m} e^{-y_i f_{t-1}(x_i)} \cdot e^{-y_i \cdot w_t \cdot h_t(x_i)}}{\sum_{i=1}^{m} e^{-y_i \cdot f_{t-1}(x_i)}} =$$

$$= \sum_{i=1}^{m}\left(\frac{e^{-y_i \cdot f_{t-1}(x_i)}}{\sum_{j=1}^{m} e^{-y_j \cdot f_{t-1}(x_j)}}\right) \cdot e^{-y_i w_t \cdot h_t(x_i)} =$$

$$\underset{\uparrow}{=} \sum_{i=1}^{m} D_i^{(t)} \cdot e^{-y_i \cdot w_t \cdot h_t(x_i)} \underset{\uparrow}{=} 2\sqrt{\epsilon_t(1-\epsilon_t)}$$

(2) סעיף   סעיף ב

(4   מתקיים $\forall t: \epsilon_t \le \frac{1}{2}-\gamma \in [0,\frac{1}{2})$

$$\Leftarrow \quad 2\sqrt{\epsilon_t(1-\epsilon_t)} \underset{\uparrow}{\le} 2\sqrt{\left(\tfrac{1}{2}-\gamma\right)\cdot\left(\tfrac{1}{2}+\gamma\right)} = \sqrt{1-4\gamma^2}$$

$g(a)=a\cdot(1-a)$ מונוטונית עולה בתחום $[0,\frac{1}{2}]$

$$\Leftarrow \quad \sqrt{1-4\gamma^2} \underset{\uparrow}{\le} \sqrt{e^{-4\gamma^2}} = e^{-\frac{4\gamma^2}{2}} = e^{-2\gamma^2}$$

$1-a \le e^{-a}$

$$\Leftarrow \quad 2\sqrt{\epsilon_t(1-\epsilon_t)} \le e^{-2\gamma^2}$$

(5   מתקיים:

$$z_0 = \frac{1}{m}\sum_{i=1}^{m} e^{-y_i F_0(x_i)} = \frac{1}{m}\sum_{i=1}^{m} e^{-0} = 1$$

טור טלסקופי

$$\prod_{t=1}^{T} \frac{Z_t}{Z_{t-1}} = \frac{Z_1}{Z_0}\cdot\frac{Z_2}{Z_1}\cdot \ldots \cdot \frac{Z_{T-1}}{Z_{T-2}} \cdot \frac{Z_T}{Z_{T-1}} = \frac{Z_T}{Z_0}$$

$$\Rightarrow Z_T = \frac{Z_T}{1} = \frac{Z_T}{Z_0} = \prod_{t=1}^{T} \frac{Z_t}{Z_{t-1}}$$

$$L_S(\hat{h}) \underset{\uparrow}{\le} Z_T = \prod_{t=1}^{T} \frac{Z_t}{Z_{t-1}} \underset{\uparrow}{\le} \prod_{t=1}^{T} e^{-2\gamma^2} = e^{-2\gamma^2 T} \quad \Leftarrow$$

(1) סעיף   מונוטוניות $\forall t: \epsilon_t \le \frac{1}{2}-\gamma$

(3)+(4) סעיפים, ומכאן ש

3ג. מתקיים כי ההפסד של מסווג האימון מקיים:

מההגדרה: $L_S(\hat{h})=\frac{1}{m}\cdot\sum_{i=1}^{m} 1\cdot[y_i\neq \hat{h}(x_i)]\in\{0,\frac{1}{m},\frac{2}{m},\dots,\frac{m-1}{m},1\}$

אם מתקיים: $L_S(\hat{h})<\frac{1}{m}$ אז בהכרח $L_S(\hat{h})=0$.

אם מתקיים: $e^{-2\gamma^2T}<\frac{1}{m}$ $\Leftarrow$ $L_S(\hat{h})=0$.

נמצא את ה-T המינימלי עבורו מתקיים הנ"ל:

$$e^{-2\gamma^2T}<\frac{1}{m}$$

$$\Rightarrow -2\gamma^2T<\log\left(\frac{1}{m}\right)=-\log(m)$$

$$\Rightarrow T>\frac{\log m}{2\gamma^2}\Rightarrow T\geq\left\lfloor 1+\frac{\log m}{2\gamma^2}\right\rfloor$$

ה.

$$\sum_{i=1}^{m} D_i^{(t+1)}\cdot 1[y_i\neq h_t(x_i)]=$$

של עדכון
מאלג׳

$$=\sum_{i=1}^{m}\frac{D_i^{(t)}e^{-w_ty_ih_t(x_i)}}{\sum_{j=1}^{m}D_j^{(t)}e^{-w_ty_jh_t(x_j)}}\cdot 1[y_i\neq h_t(x_i)]=$$

מסעיף ב'

$$=\frac{\sum_{i=1}^{m}D_i^{(t)}e^{-w_ty_ih_t(x_i)}\cdot 1[y_i\neq h_t(x_i)]}{2\sqrt{\epsilon_t(1-\epsilon_t)}}=$$

$y_i=h_t(x_i)\Rightarrow$
$1[y_i\neq h_t(x_i)]=0$

$$=\frac{\sum_{i=1,\ y_i\neq h_t(x_i)}^{m}D_i^{(t)}\cdot e^{w_t}}{2\sqrt{\epsilon_t(1-\epsilon_t)}}=\frac{\epsilon_t e^{w_t}}{2\sqrt{\epsilon_t(1-\epsilon_t)}}=$$

של
עדכון ϵ_t

של עדכון
w_t

$$=\frac{\epsilon_t\cdot e^{\frac{1}{2}\log\left(\frac{1}{\epsilon_t}-1\right)}}{2\sqrt{\epsilon_t(1-\epsilon_t)}}=\frac{\epsilon_t\sqrt{\frac{1-\epsilon_t}{\epsilon_t}}}{2\sqrt{\epsilon_t(1-\epsilon_t)}}=$$

$$=\frac{\sqrt{\epsilon_t(1-\epsilon_t)}}{2\sqrt{\epsilon_t(1-\epsilon_t)}}=\frac{1}{2}$$

■

Qustion 2:

In [25]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import sklearn.datasets
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
```

In [5]:
```
df = pd.read_csv('winequality-red.csv', delimiter=';')
df['quality'] = (df['quality'] > 5).astype(int)
X, y = df.drop(columns=['quality']).to_numpy(), df['quality'].to_numpy()
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.40, random_state = 42)
```

1.

In [18]:
```
dt = DecisionTreeClassifier(max_depth=12, random_state=0)
dt = dt.fit(X_train, y_train)
train_accuracy = accuracy_score(y_true=y_train, y_pred=dt.predict(X_train))
test_accuracy = accuracy_score(y_true=y_test, y_pred=dt.predict(X_test))
print(f"Train Accuracy: {train_accuracy*100:.2f}%")
print(f"Test Accuracy: {test_accuracy*100:.2f}%")
```

```
Train Accuracy: 98.96%
Test Accuracy: 72.81%
```

2.

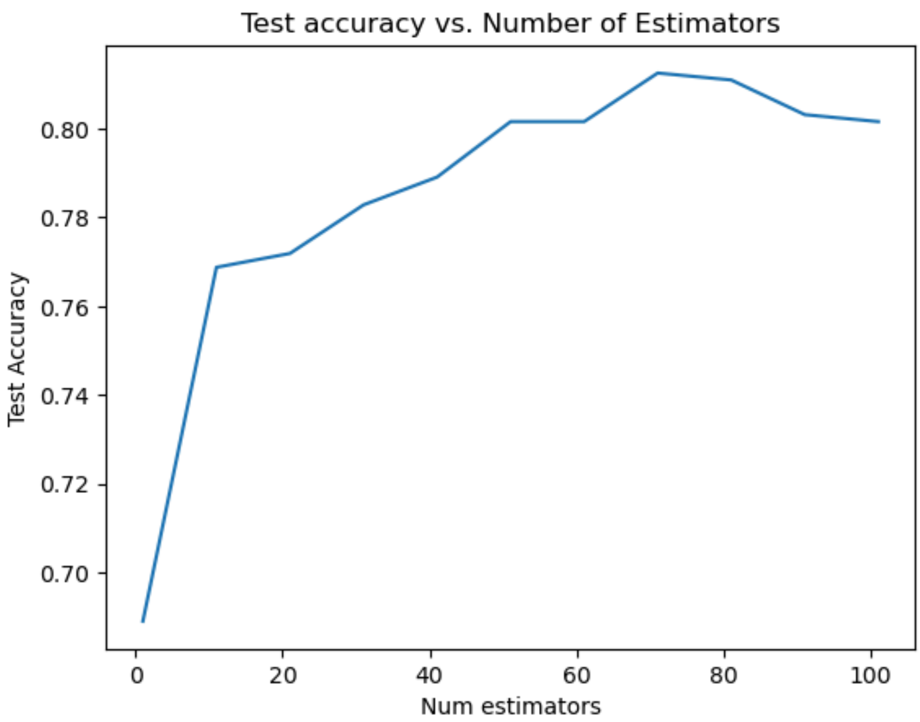
In [38]:
```
rf = RandomForestClassifier(n_estimators=100, max_depth=12, random_state=0)
rf = rf.fit(X_train, y_train)
train_accuracy = accuracy_score(y_true=y_train, y_pred=dt.predict(X_train))
test_accuracy = accuracy_score(y_true=y_test, y_pred=dt.predict(X_test))
print(f"Train Accuracy: {train_accuracy*100:.2f}%")
print(f"Test Accuracy: {test_accuracy*100:.2f}%")

values = []
n_estimators = list(range(1, 110, 10))

for n in n_estimators:
    rf = RandomForestClassifier(n_estimators=n, max_depth=12, random_state=0)
    rf = rf.fit(X_train, y_train)
    values.append(accuracy_score(y_true=y_test, y_pred=rf.predict(X_test)))

plt.plot(n_estimators, values)
plt.title('Test accuracy vs. Number of Estimators')
plt.xlabel('Num estimators')
plt.ylabel('Test Accuracy')
plt.show()
```

```
Train Accuracy: 99.90%
Test Accuracy: 80.16%
```



3.

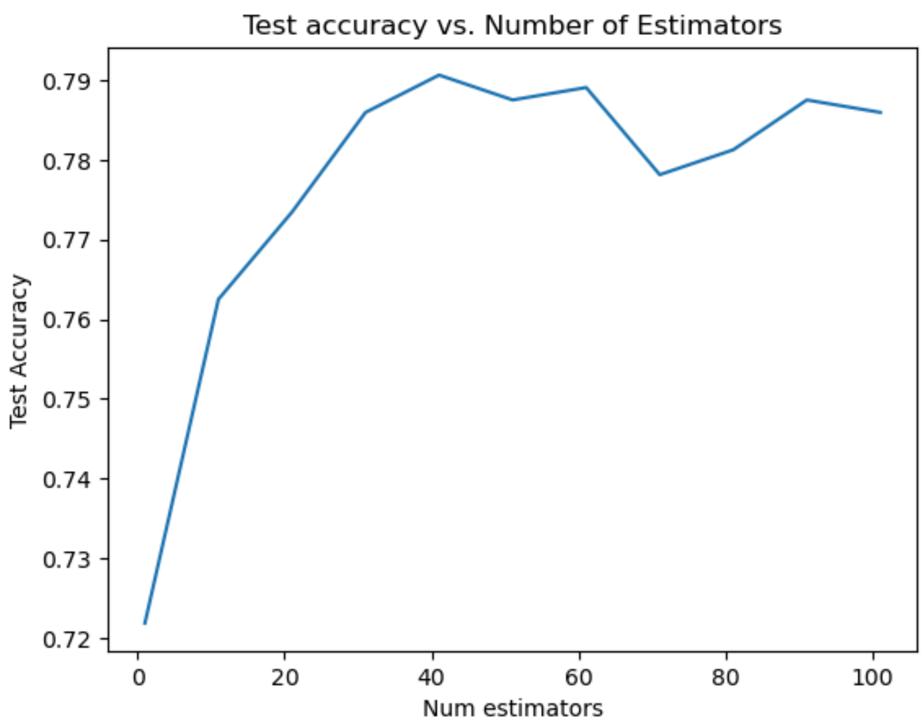
In [47]:
```
rf = RandomForestClassifier(n_estimators=100, max_depth=12, random_state=0, max_features=11)
rf = rf.fit(X_train, y_train)
train_accuracy = accuracy_score(y_true=y_train, y_pred=rf.predict(X_train))
test_accuracy = accuracy_score(y_true=y_test, y_pred=rf.predict(X_test))
print(f"Train Accuracy: {train_accuracy*100:.2f}%")
print(f"Test Accuracy: {test_accuracy*100:.2f}%")

values = []
n_estimators = list(range(1, 110, 10))

for n in n_estimators:
    rf = RandomForestClassifier(n_estimators=n, max_depth=12, random_state=0, max_features=1
    rf = rf.fit(X_train, y_train)
    values.append(accuracy_score(y_true=y_test, y_pred=rf.predict(X_test)))

plt.plot(n_estimators, values)
plt.title('Test accuracy vs. Number of Estimators')
plt.xlabel('Num estimators')
plt.ylabel('Test Accuracy')
plt.show()
```

```
Train Accuracy: 99.90%
Test Accuracy: 78.75%
```



4.

A) The model in section B is better than the model in section A because it has more regularization. Instead of relying on a single tree, we use the principle of bootstrapping and employ 100 trees.

B) The model in section C is worse than the model in section B because we don't prevent it from overfitting. The model uses all the features at every split, instead of just the square root of the number of features, which is the default value in the library.

5. a. נסיק כי הפיצול הטוב יותר הוא פיצול 2

כבר בהתפלגות הרשומה קיבלנו זאת (בנוסף החישוב ניתן לראות שמדד ה- impurity עבור קודקוד ימני הוא אפס בכל המדדים.

למרות שהפיצול הראשון מחזיר עלה יותר מאוזן, הפיצול השני מחזיר עלה עם מדדי impurity נמוכים יותר ולכן הוא עדיף.

6. פיצול $1 = S_1$:

$$P_L^1\left(\frac{150}{200}=\frac{3}{4},\ \frac{50}{200}=\frac{1}{4}\right),\quad P_R^1=\left(\frac{1}{4},\frac{3}{4}\right)$$

Entropy:

$$R_m^{left} = H_{Entropy}(R_m^{left}) = -\frac{1}{4}\log\left(\frac{1}{4}\right)-\frac{3}{4}\cdot\log\left(\frac{3}{4}\right) = 0.244$$

$$R_m^{right} = H_{Entropy}(R_m^{right}) = -\frac{1}{4}\log\left(\frac{1}{4}\right)-\frac{3}{4}\cdot\log\left(\frac{3}{4}\right) = 0.244$$

$$G(R_m, S_1) = \frac{1}{2}\cdot 0.244 + \frac{1}{2}\cdot 0.244 = 0.244$$

Gini :

$$R_m^{left} = H_{Gini}(R_m^{left}) = 1-\left(\frac{1}{4}\right)^2-\left(\frac{3}{4}\right)^2 = \frac{3}{8}$$

$$R_m^{right} = H_{Gini}(R_m^{right}) = 1-\left(\frac{1}{4}\right)^2-\left(\frac{3}{4}\right)^2 = \frac{3}{8}$$

$$G(R_m, S_1) = \frac{1}{2}\cdot\frac{3}{8}+\frac{1}{2}\cdot\frac{3}{8} = \frac{3}{8} = 0.375$$

Misclassification:

$$R_m^{left} = H_{M.Error}(R_m^{left}) = 1-\max\left\{\frac{1}{4},\frac{3}{4}\right\} = \frac{1}{4}$$

$$R_m^{right} = H_{M.Error}(R_m^{right}) = 1-\max\left\{\frac{1}{4},\frac{3}{4}\right\} = \frac{1}{4}$$

$$G(R_m, S_1) = \frac{1}{2}\cdot\frac{1}{4}+\frac{1}{2}\cdot\frac{1}{4} = \frac{1}{4} = 0.25$$

פיצול 2: S_2 :

$$P_L^2 = \left(\frac{100}{300} = \frac{1}{3}, \frac{200}{300} = \frac{2}{3}\right), \quad P_R^2 = \left(\frac{100}{100} = 1, \frac{0}{100} = 0\right)$$

Entropy:
$$R_m^{left} = H_{Entropy}(R_m^{left}) =$$
$$= -\frac{1}{3}\log\left(\frac{1}{3}\right) - \frac{2}{3}\cdot\log\left(\frac{2}{3}\right) = 0.276$$
$$R_m^{right} = H_{Entropy}(R_m^{right}) =$$
$$= -1\log(1) - 0\cdot\log(0) = 0$$
$$G(R_m, S_2) = \frac{1}{4}\cdot 0 + \frac{3}{4}\cdot 0.276 = 0.138$$

Gini:
$$R_m^{left} = H_{Gini}(R_m^{left}) =$$
$$= 2\cdot\frac{1}{3}\cdot\frac{2}{3} = \frac{4}{9}$$
$$R_m^{right} = H_{Gini}(R_m^{right}) =$$
$$= 2\cdot 1\cdot 0 = 0$$
$$G(R_m, S_2) = \frac{1}{4}\cdot 0 + \frac{3}{2}\cdot\frac{4}{9} = \frac{1}{3} = 0.333$$

Misclassification:
$$R_m^{left} = H_{M.Error}(R_m^{left}) =$$
$$= 1 - \max\left\{\frac{2}{3}, \frac{1}{3}\right\} = \frac{1}{3}$$
$$R_m^{right} = H_{M.Error}(R_m^{right}) =$$
$$= 1 - \max\{1, 0\} = 0$$
$$G(R_m, S_2) = \frac{3}{4}\cdot\frac{1}{3} + \frac{1}{4}\cdot 0 = \frac{1}{4} = 0.25$$

המדד שאינו מוצלח לבחירת הפיצול הוא Misclassification Rate.
שני המדדים האחרים מתחשבים בהתפלגות של הקלאסים
בעוד ש- M.Rate מתחשב רק בקלאס המקסימלי בקודקוד.
כתוצאה מכך, אין ביטוי ל"ערבוב" הקלאסים בקודקוד.
כפי שראינו בחישובים, המדד החזיר תוצאה זהה עבור שני
פיצולים שונים למדי.