

Основне академске студије  
Информациони инжењеринг

Методе и технике науке о подацима

# Класификација

(помоћни материјали)

1. Увод у класификацију
2. Извори и литература

## Класификација

разврставање појава по класама

задатак класификације

задатак у којем треба за посматране појаве проценити којим од могућих класа припадају

појединачној посматраној појави бива додељена ознака (лабела) која одговара процењеној класи

стварна и процењена класа за појединачну појаву могу се подударити али и не морају

врсте класификације

бинарна класификација

скуп могућих класа обухвата две класе

$n$ -арна класификација

скуп могућих класа обухвата  $n$  класа, где је  $n > 2$

## Класификација

процена припадности класи начелно се одвија на основу вредности одабраних обележја

**циљно обележје** (зависно обележје, одговор)

обележје чије су вредности ознаке (лабеле) класа

вредност циљног обележја треба одредити на основу вредности других обележја по врсти је категоријско (квалитативно)

обично постоји само једно циљно обележје у класификацији  
обележје  $Y$

**предикторско обележје** (независно обележје, предиктор)

обележје чије се вредности користе при одређивању вредности циљног обележја по врсти је категоријско (квалитативно) или континуално (квантитативно)

обично постоји више предикторских обележја у класификацији

обележја  $X_1, X_2, \dots, X_p$

$p$  је број предикторских обележја

## Класификациони модел

класификациони модел је модел који начелно повезује циљно обележје и предикторска обележја тако да је могуће проценити вредност циљног обележја

примена надгледаног учења у класификацији

обучавање класификационог модела

припрема класификационог модела за потребе обављања задатка класификације

обучавање се изводи на основу података који обухватају циљно обележје и предикторска обележја

## Методи класификације

разноврсни методи могу бити примењени у циљу класификације

регресија

логистичка регресија

метод најближих суседа

метод потпорних вектора

стабла одлучивања

појединачно стабло одлучивања

вишеструка стабла одлучивања

неуронске мреже

...

## Оцењивање перформанси у класификацији

### грешка обучавања

процент појава које су коришћене током обучавања а применом класификационог модела погрешно су класификоване

### грешка тестирања

процент појава које нису коришћене током обучавања а применом класификационог модела погрешно су класификоване

## Оцењивање перформанси у класификацији

### класични валидациони приступ

расположиви подаци о појавама бивају подељени на скуп за обучавање и валидациони скуп

класификациони модел бива подешаван на основу података из скупа за обучавање перформансе подешеног класификационог модела бивају процењене над подацима из валидационог скупа

рачунање грешке над валидационим скупом

расположиви подаци могу бити подељени у разним односима

нпр. подела у односу 75% (скуп за обучавање) наспрам 25% (валидациони скуп)

## Оцењивање перформанси у класификацији

### матрица конфузије

представа перформанси класификационог модела по могућим класама  
пример за случај бинарне класификације (могуће класе 0 и 1)

број стварно позитивних (енгл. *true positive*)

број исправно класификованих појава класе 1

број стварно негативних (енгл. *true negative*)

број исправно класификованих појава класе 0

број лажно позитивних (енгл. *false positive*)

број неисправно класификованих појава класе 0

број лажно негативних (енгл. *false negative*)

број неисправно класификованих појава класе 1

|                 |   | стварна класа                |                              |
|-----------------|---|------------------------------|------------------------------|
|                 |   | 0                            | 1                            |
| процењена класа | 0 | број стварно негативних (TN) | број лажно негативних (FN)   |
|                 | 1 | број лажно позитивних (FP)   | број стварно позитивних (TP) |

## Оцењивање перформанси у класификацији

### матрица конфузије

представа перформанси класификационог модела по могућим класама  
пример за случај бинарне класификације (могуће класе 0 и 1)

тачност (енгл. *accuracy*)

$$acc = \frac{TP + TN}{TP + TN + FP + FN}$$

прецизност (енгл. *precision*)

$$pre = \frac{TP}{TP + FP}$$

осетљивост (енгл. *sensitivity*)

$$sen = \frac{TP}{TP + FN}$$

специфичност (енгл. *specificity*)

$$spe = \frac{TN}{TN + FP}$$

скор *F1* (енгл. *F1 score*)    **процењена  
класа**

$$F1 = \frac{2}{\frac{1}{pre} + \frac{1}{sen}}$$

|                    |   | стварна<br>класа                            |   |
|--------------------|---|---|---|
|                    |   | 0   | 1   |
| процењена<br>класа | 0 | број стварно<br>негативних<br>( <i>TN</i> ) | број лажно<br>негативних<br>( <i>FN</i> )   |
|                    | 1 | број лажно<br>позитивних<br>( <i>FP</i> )   | број стварно<br>позитивних<br>( <i>TP</i> ) |

1. Увод у класификацију
- 2. Извори и литература**

## Основни извори и литература

- ◆ James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: With Applications in R. Springer (New York, NY, USA); 2013.
- ◆ Shalev-Shwartz S, Ben-David S. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press (New York, NY, USA); 2014.
- ◆ Ault SV, Liao SN, Musolino L. Principles of Data Science. OpenStax, Rice University (Houston, TX, USA); 2025. Internet: <https://openstax.org/details/books/principles-data-science/>

Основне академске студије  
Информациони инжењеринг

Методе и технике науке о подацима

# Класификација

(помоћни материјали)