Machine learning (ML) elmélet

Tartalomjegyzék

[Lesson 1 – PreProcessing data 3](#_Toc2195812)

[Lépések 3](#_Toc2195813)

[Dataset 3](#_Toc2195814)

[Missing data 3](#_Toc2195815)

[Categorical variable 3](#_Toc2195816)

[Training sets and testing sets 3](#_Toc2195817)

[Standardizáció és normalizálás 3](#_Toc2195818)

[Lesson 2 – Simple Linear Regression 4](#_Toc2195819)

[Képlet 4](#_Toc2195820)

[Közönséges legkisebb négyzetek: 4](#_Toc2195821)

# Lesson 1 – PreProcessing data

## Lépések

Rövid leírások a témakörben található fogalmakról

### Dataset

Elemzendő adathalmaz

Tulajdonságai:

* tartalma releváns
* egy vagy több összefüggő változó (dependant variable DV)
* további adatok, amikről feltételezzük, hogy függőségben vannak a függő változókkal

Cél: Ezen feltételezéseket bizonyítani, majd ezen függőségek alapján megjósolni a függő változók alakulását

### Missing data

Előfeltétel, hogy minden elemi adatsor teljes legyen.

Pl: Medián, Átlag, Leggyakoribb érték

### Categorical variable

A dataset egy része mérhető adatokból áll, másik része viszont kategorizálható adatokból

Cél: olyan számszerű adatokat rendelünk az adott változókhoz, amelyet ennél fogva képes megemészteni az ML model.

* IGAZ/HAMIS
* elemi IGAZ/HAMIS: Annyi oszlopra kell felbontani, amennyi lehetséges értéket felvehet. 🡪 A legtöbb ML algoritmus egy számszerű skála esetén súlyozza a magasabb értékeket és így eltorzulhat a feltételezés.

Például:

|  |  |
| --- | --- |
| Magyarország | 1 |
| Németország | 6 |

|  |  |  |
| --- | --- | --- |
| Magyarország | 1 | 0 |
| Németország | 0 | 1 |

VAGY

### Training sets and testing sets

A dataset-et fel kell bontani egy tanuló (train) és egy tesztelési (test) darabra.

### Standardizáció és normalizálás

Releváns lehet az értékek közti távolság pl.: életkor ~ bér

* Euklédeszi távolság: P1 és P2 pont euklédeszi távolsága

#### Standardizáció:

#### Normalizáció:

## Lesson 2 – Simple Linear Regression

### Képlet

Dependant Constant együttható independant

variable (DV) (coefficient) variable (IV)

Salary ($)

+10k

Constant

30k

1 year

Experience

### Közönséges legkisebb négyzetek: