

Pattern recognition - 4. lab

Feature selection

Viktor Kocur
viktor.kocur@fmph.uniba.sk

DAI FMFI UK

10.3.2020

Entropy

Shannon entropy

$$H(Y) = - \sum_{y \in \omega} P(Y = y) \cdot \log_2(P(Y = y))$$

X	Y
Matematika	Áno
História	Nie
Informatika	Áno
Matematika	Nie
Matematika	Nie
Informatika	Áno
História	Nie
Matematika	Áno

Entropy

Shannon entropy

$$H(Y) = - \sum_{y \in \omega} P(Y = y) \cdot \log_2(P(Y = y))$$

X	Y
Matematika	Áno
História	Nie
Informatika	Áno
Matematika	Nie
Matematika	Nie
Informatika	Áno
História	Nie
Matematika	Áno

■ $H(X) = 1.5$

Entropy

Shannon entropy

$$H(Y) = - \sum_{y \in \omega} P(Y = y) \cdot \log_2(P(Y = y))$$

X	Y
Matematika	Áno
História	Nie
Informatika	Áno
Matematika	Nie
Matematika	Nie
Informatika	Áno
História	Nie
Matematika	Áno

■ $H(X) = 1.5$

■ $H(Y) = 1$

Entropy

Specific conditional entropy

$H(Y|X = v) = H(Y)$, only for values of Y , where $X = x$

X	Y
Matematika	Áno
História	Nie
Informatika	Áno
Matematika	Nie
Matematika	Nie
Informatika	Áno
História	Nie
Matematika	Áno

Entropy

Specific conditional entropy

$H(Y|X = v) = H(Y)$, only for values of Y , where $X = x$

X	Y
Matematika	Áno
História	Nie
Informatika	Áno
Matematika	Nie
Matematika	Nie
Informatika	Áno
História	Nie
Matematika	Áno

■ $H(Y|X = M) = 1$

Entropy

Specific conditional entropy

$H(Y|X = v) = H(Y)$, only for values of Y , where $X = x$

X	Y
Matematika	Áno
História	Nie
Informatika	Áno
Matematika	Nie
Matematika	Nie
Informatika	Áno
História	Nie
Matematika	Áno

■ $H(Y|X = M) = 1$

■ $H(Y|X = H) = 0$

Entropy

Specific conditional entropy

$H(Y|X = v) = H(Y)$, only for values of Y , where $X = x$

X	Y
Matematika	Áno
História	Nie
Informatika	Áno
Matematika	Nie
Matematika	Nie
Informatika	Áno
História	Nie
Matematika	Áno

- $H(Y|X = M) = 1$
- $H(Y|X = H) = 0$
- $H(Y|X = I) = 0$

Entropy

Conditional entropy

$$H(Y|X) = \sum_{x \in \Omega_x} P(X = x) \cdot H(Y|X = x)$$

X	Y
Matematika	Áno
História	Nie
Informatika	Áno
Matematika	Nie
Matematika	Nie
Informatika	Áno
História	Nie
Matematika	Áno

Entropy

Conditional entropy

$$H(Y|X) = \sum_{x \in \Omega_x} P(X = x) \cdot H(Y|X = x)$$

X	Y
Matematika	Áno
História	Nie
Informatika	Áno
Matematika	Nie
Matematika	Nie
Informatika	Áno
História	Nie
Matematika	Áno

■ $H(Y|X) = 0.5$

Entropy - Matlab

Load the data

```
load census1994  
Y = categorical(adultdata.salary);  
X1 = categorical(adultdata.education);
```

Exercise

Create a function `entropia(X)`, which returns the value we defined as $H(X)$.

countcats

`countcats(X)` - returns the number of elements in each category of the categorical array `X`.

Entropy - Matlab

Exercise

Create a function `podm_entropia(Y, X)`, which returns the value we defined as $H(Y|X)$.

categories

`c = categories(X)` - returns a cell array of character vectors containing the categories of the categorical array `X`. Note: to get a category we need to use notation `c{i}`, we can use this when indexing by a logical matrix (`X == c{i}`).

vráti cell štruktúru `c` s jednolivými kategóriami z `X` (ak je `X` typu categorical). Pozn.: z cell dostane kategóriu ako `c{i}` a použijeme pri indexácii logickou maticou (`X == c{i}`).