

Aula11

DATA SCIENCE IPT

TURMA 02

Teorema de Bayes ampliado ($p(y|x_1, x_2 \dots x_n)$):

$$P(y | x_1, \dots, x_n) = \frac{P(y)P(x_1, \dots, x_n | y)}{P(x_1, \dots, x_n)}$$

A hipótese simplificadora....independência entre features:

$$P(x_i | y, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = P(x_i | y).$$

A probabilidade de uma classe, dadas n features, pela simplificação:

$$P(y | x_1, \dots, x_n) = \frac{P(y) \prod_{i=1}^n P(x_i | y)}{P(x_1, \dots, x_n)}$$

A probabilidade de cada classe é proporcional a um produto:

$$P(y \mid x_1, \dots, x_n) \propto P(y) \prod_{i=1}^n P(x_i \mid y)$$

↓

$$\hat{y} = \arg \max_y P(y) \prod_{i=1}^n P(x_i \mid y),$$

A classe prevista é a de maior “produto”, já que o denominador $P(x_1, x_2, \dots, x_n)$ é o mesmo.

Assim, as diferentes versões do algoritmo Naive Bayes, envolvem maneiras diferentes de calcular $p(x_i | y)$...

No caso de Naïve Bayes Gaussiano:

$$P(x_i | y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$

Ver notebook gaussian bayes_res.ipynb

NLP: o problema de extrair features de textos

Bag of Words

Intelligent applications creates intelligent business processes

intelligent	applications	creates	business	processes
2	1	1	1	1

Peso de uma palavra
com tf*idf

$$w_{i,j} = tf_{i,j} \times \log \left(\frac{N}{df_i} \right)$$

$tf_{i,j}$ = number of occurrences of i in j
 df_i = number of documents containing i
 N = total number of documents

Coming soon

Vamos criar um protótipo de chatbot simplificado.

As frases digitadas serão, por multiclassificação, associadas a “intents”, mas o diálogo não passará disso...

A ideia é atender pessoas que querem saber informações sobre um carro
(preço, performance, consumo, modelos, garantia)

Parte 1: analisar notebook chatbot_res.ipynb

Quando o diálogo começa, o contexto é “geral”, imagine que o usuário perguntou sobre “modelos”...indo para esse novo contexto. Como seria o modelo nesse contexto?

Partindo de chatbot_res.ipynb, crie modelos.ipynb e responda sobre modelos...

Veja na documentação do sklearn como é feito o cálculo de idf

Prof. Ademar Sigrist



Cursos com Alta Performance de
Aprendizado

© 2019 – Linked Education