

Procesamiento de Lenguaje Natural

Eric S. Téllez

Mario Graff

Tabla de contenidos

Prefacio	3
Notación	4
1 Introducción	5
2 Manejando Texto	6
3 Modelado de Lenguaje	7
4 Clasificación de Texto	8
5 Representación de Texto	9
6 Mezcla de Modelos	10
7 Tareas de Clasificación de Texto	11
8 Bases de Conocimiento	12
9 Visualización	13
10 Conclusiones	14
Referencias	15

Prefacio

El curso trata de ser auto-contenido, es decir, no debería de ser necesario leer otras fuentes para poder entenderlo y realizar las actividades. De cualquier manera es importante comentar que el curso está basado en los siguientes libros de texto:

- Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Third Edition draft. Daniel Jurafsky and James H. Martin. [pdf](#)
- Introduction to machine learning, Third Edition. Ethem Alpaydin. MIT Press.
- An Introduction to Statistical Learning with Applications in R. Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. Springer Texts in Statistics.
- All of Statistics. A Concise Course in Statistical Inference. Larry Wasserman. MIT Press.
- An Introduction to the Bootstrap. Bradley Efron and Robert J. Tibshirani. Monographs on Statistics and Applied Probability 57. Springer-Science+Business Media.
- Understanding Machine Learning: From Theory to Algorithms. Shai Shalev-Shwartz and Shai Ben-David. Cambridge University Press.

Notación

La Tabla 1 muestra la notación que se seguirá en este documento.

Tabla 1: Notación

Símbolo	Significado
x	Variable usada comunmente como entrada
y	Variable usada comunmente como salida
\mathbb{R}	Números reales
\mathbf{x}	Vector Columna $\mathbf{x} \in \mathbb{R}^d$
d	Dimensión
$\mathbf{w} \cdot \mathbf{x}$	Producto punto donde \mathbf{w} y $\mathbf{x} \in \mathbb{R}^d$
\mathcal{D}	Conjunto de datos
\mathcal{T}	Conjunto de entrenamiento
\mathcal{V}	Conjunto de validación
\mathcal{G}	Conjunto de prueba
N	Número de ejemplos
K	Número de clases
$\mathbb{P}(\cdot)$	Probabilidad
\mathcal{X}, \mathcal{Y}	Variables aleatorias
$\mathcal{N}(\mu, \sigma^2)$	Distribución Normal con parámetros μ y σ^2
$f_{\mathcal{X}}$	Función de densidad de probabilidad de \mathcal{X}
$\mathbb{1}(e)$	Función para indicar; 1 only if e is true
Ω	Espacio de búsqueda
\mathbb{V}	Varianza
\mathbb{E}	Esperanza

1 Introducción

El **objetivo** de la unidad es

2 Manejando Texto

El **objetivo** de la unidad es

3 Modelado de Lenguaje

El **objetivo** de la unidad es

4 Clasificación de Texto

El **objetivo** de la unidad es

5 Representación de Texto

El **objetivo** de la unidad es

6 Mezcla de Modelos

El **objetivo** de la unidad es

7 Tareas de Clasificación de Texto

El **objetivo** de la unidad es

8 Bases de Conocimiento

El **objetivo** de la unidad es

9 Visualización

El **objetivo** de la unidad es

10 Conclusiones

El **objetivo** de la unidad es

Referencias