

MULTIMODAL MACHINE LEARNING

Prof. Francesco Camastra

Obiettivi

Lo scopo del corso è fornire i fondamenti teorici e pratici per l'elaborazione di dati multimodali, quali l'audio e il video mediante metodi di Machine Learning.

Conoscenza e capacità di comprensione

Lo studente deve dimostrare di conoscere e saper utilizzare i fondamenti dei più moderni linguaggi di programmazione, quali ad esempio quelli Orientati agli Oggetti (OO). Lo studente dovrà avere sufficiente familiarità con le tecniche di algebra lineare, analisi matematica, probabilità e statistica, impartite nel corso di laurea triennale, in modo da poter comprendere i fondamenti teorici delle tecniche di machine learning presentate nel corso. Infine, lo studente dovrà avere una buona dimestichezza con la lingua inglese scritta in modo tale da poter avvalersi delle risorse bibliografiche, sotto forma di testi e di articoli scientifici, per la quasi totalità in lingua inglese.

Capacità applicative

Lo studente deve dimostrare di saper utilizzare la propria conoscenza acquisita per la risoluzione di un problema di classificazione o di previsione di dati multimodali utilizzando tecniche di Machine learning. Deve inoltre dimostrare capacità di sviluppo e di analisi di complessità di algoritmi. Lo studente dovrà dimostrare un uso consapevole di librerie software, attualmente disponibili, per l'implementazione dei principali metodi di machine learning presentati nel corso.

Autonomia di giudizio

Lo studente deve essere in grado di sapere valutare in maniera autonoma l'efficacia e l'efficienza di un'applicazione di machine learning ad un caso reale. Dovrà aver, inoltre, maturato una sufficiente indipendenza e senso critico in modo da poter utilizzare le risorse bibliografiche. Tale capacità è cruciale per la maturazione di attitudini sia di self-learning, durante il corso degli studi, che di Lifelong learning, che lo studente magistrale dovrà essere in grado di svolgere, dopo aver completato gli studi magistrali.

Abilità nella comunicazione

Lo studente deve essere in grado di redigere una relazione e di organizzare una presentazione, entrambe preferibilmente in lingua inglese, su un'applicazione di machine learning multimodale.

Capacità di apprendere

Lo studente dovrà aver maturato, alla fine del corso, la capacità di Self-Learning, ossia di essere in grado, ai fini dell'autoapprendimento, di utilizzare, in modo indipendente, le risorse bibliografiche che motori di ricerca e repository open source, quali Scholar Google e Researchgate, mettono a disposizione.

Prerequisiti

Nozioni di base di Programmazione Orientata agli Oggetti, di Algebra Lineare, Analisi Matematica, Probabilità e Statistica e di Calcolo Scientifico.

Contenuti

Il corso intende fornire i fondamenti teorici e pratici per l'elaborazione di dati multimodali tramite tecniche di Machine Learning.

Programma esteso

Programma

Multimodal Machine Learning

1. Introduction to Image and Video Processing
 - 1.1 Human Color Perception
 - 1.2 Color Models: Hardware-Oriented Models
 - 1.3 Color Models: Perceptually Uniform Color Models
 - 1.4 Image Standards: JPEG and JPEG 2000
 - 1.5 Wavelet Transform
 - 1.6 MPEG Standards: MPEG-2 and MPEG-4

2. Audio Processing
 - 2.1 Human Sound Perception
 - 2.2 Speech Production
 - 2.3 Speech Feature Extraction
 - 2.4 Speech Classifiers

3. Complements of Machine Learning
 - 3.1 Feature Selection Methods
 - 3.2 Manifold Learning Methods
 - 3.3 Kernel and Spectral Methods
 - 3.4 Sequence Classification Methods: HMM and Viterbi Algorithm

4. MultiModal Recognition by Machine Learning
 - 4.1 Automatic Face Recognition
 - 4.2 Video Segmentation and Keyframe Extraction
 - 4.3 Gesture Recognition
 - 4.4 Speech Recognition
 - 4.5 Automatic Personality Perception

Metodi Didattici

La didattica è svolta mediante lezioni frontali, e da parte degli stessi studenti su alcune tematiche di loro particolare interesse. Sulla piattaforma di riferimento del corso c'è la possibilità di usufruire di video-lezioni.

Verifica dell'apprendimento

Lo studente deve avere dimostrato la capacità di comprensione di tecniche di machine learning per l'analisi di dati multimediali. Per il superamento dell'esame lo studente dovrà sviluppare un progetto ed effettuare un colloquio orale. Il progetto, concordato con il docente, riguarderà applicazione di tecniche di machine learning ad uno specifico dominio multimodale. Lo sviluppo del progetto sarà corredato dalla redazione di una relazione, preferibilmente in lingua inglese. Nel colloquio orale lo studente presenterà e discuterà il progetto effettuato e dovrà mostrare di aver appreso i metodi di machine learning trattate durante il corso.

Testi

- T. Hastie, R. Tibshirani, J. Friedman, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", 2nd Edition, 2008, Springer, ISBN: 978-0387848570
- M. Gori, "Machine Learning: A Constraint-Based Approach", 2017, Morgan Kaufman, ISBN: 978-0081006597
- F. Camastra, A. Vinciarelli, "Machine Learning for Audio, Image and Video Analysis: Theory and Applications", 2nd Edition, 2016, Springer Verlag, ISBN: 978-1447168409
- J. Shawe-Taylor, N. Cristianini, "Kernel Methods for Pattern Analysis", 2004, Cambridge University Press, ISBN: 978-0521813976