

Le basi del Machine Learning

Trasformare enormi dati di volumi in informazioni utili

 A distanza

 3 giorni (21 Ore)

Open : 1.990,00 € +IVA

WebCode: IT.29

Packaged in azienda : 5.380,00 € +IVA +10% di
Project Management (Quota riferita ad un gruppo
di 10 pax max)

Customized : Su richiesta

La padronanza del Data Mining e del Machine Learning è diventata un'abilità necessaria, addirittura **essenziale** per chiunque desideri sviluppare competenze sui Big Data poiché consente loro di esplorare o cercare volumi di dati molto grandi per costruire **modelli** e rispondere a un'ampia varietà di **problemi di business** quando i metodi statistici tradizionali diventano inefficaci. Per questo, gli esperti di Big Data devono padroneggiare lo sviluppo e lo studio di **algoritmi che consentano alle macchine di apprendere automaticamente dai dati** ed eseguire attività in modo autonomo per modellare le tendenze.

A chi è rivolto



Per chi

- Ingegneri, analisti, responsabili marketing
- Data Analyst, Data Scientist, Data Steward
- Chiunque sia interessato alle tecniche di Data Mining e Machine Learning



Prerequisiti

Conoscere l'utilità del Data Mining e le problematiche dei Big Data nel targeting economico

Programma

1 - Apprendimento automatico - Introduzione

- Introduzione
- Aree di competenza
- Concentrarsi sulla scienza dei dati (data mining)
- Concentrarsi sull'apprendimento automatico
- Concentrarsi sui Big Data
- Concentrarsi sull'apprendimento profondo
- Definizione di apprendimento automatico
- Esempi di attività di apprendimento automatico
- Cosa possono imparare le macchine
- Le diverse modalità di allenamento

2 - I fondamenti dell'apprendimento automatico

- Preambolo: - Un problema di ottimizzazione - Ricerca della capacità ottimale del modello - Capacità di relazione ed errori - Un contributo filosofico - Quadro statistico - Anatomia di un modello di machine learning
- Set di dati di addestramento: - Quadro statistico - Variabili predittive - Catena di elaborazione delle variabili predittive - Variabili da prevedere

- Funzioni di ipotesi: - Principio: insiemi di funzioni di ipotesi - Contesto di selezione delle funzioni di ipotesi - Caratteristiche delle funzioni di ipotesi - Modelli probabilistici frequentisti e bayesiani
- Funzioni di costo: - Stimatori - Principio di massima verosimiglianza (MLE*) - MAP - Massimo A Posteriori - Il bias di uno stimatore - La varianza di uno stimatore - Il compromesso bias-varianza - Funzioni di costo - La regolarizzazione dei parametri
- Algoritmi di ottimizzazione: - Le principali classi di algoritmi di ottimizzazione - Discesa del gradiente (1° ordine) - Discesa del gradiente (dettagli) - Approcci di Newton (2° ordine) - Ottimizzazione batch e stocastica - Per approfondire

Laboratorio: Implementazione dell'ambiente di lavoro di Machine Learning

3 - Classificazione

- Introduzione: - Scelta di un algoritmo di classificazione
- Regressione logistica: - Da Perceptron alla regressione logistica - Presupposti del modello - Apprendimento dei pesi del modello - Esempio di implementazione: scikit-learn - Regressione logistica - Scheda riassuntiva
- SVM: - Classificazione con margine massimo - La nozione di margine morbido - Macchine del kernel - Trucco del kernel - Funzioni del kernel - SVM - Matematica - SVM - Foglio riassuntivo
- Alberi decisionali: - Principio di base - Funzionamento - Massimizzazione del guadagno informativo - Misurazione dell'impurità di un nodo - Esempio di implementazione: scikit-learn - Alberi decisionali - Foglio riassuntivo
- K vicini più vicini (kNN): - Apprendimento basato su esempi - Principio di funzionamento - Vantaggi e svantaggi - kNN - Scheda riassuntiva
- Sintesi

Laboratorio: Sperimentazione di algoritmi di classificazione su casi concreti

4 - Pratiche

- Preelaborazione: - Gestione dei dati mancanti - Trasformatori e stimatori - Elaborazione dei dati categorici - Partizionamento dei set di dati - Ridimensionamento dei dati
- Ingegneria delle variabili predittive (Feature Engineering): - Selezione delle variabili predittive - Selezione indotta dalla regolarizzazione L1 - Selezione sequenziale delle variabili - Determinazione dell'importanza delle variabili - Riduzione dimensionale mediante compressione dei dati - Estrazione delle variabili predittive - Component analysis (PCA) - Lineare Analisi discriminante (LDA) - Kernel PCA (KPCA)
- Impostazioni degli iperparametri e valutazione del modello: - Buone pratiche - La nozione di Pipeline - Convalida incrociata (convalida incrociata) - Curve di apprendimento - Curve di convalida - Ricerca in griglia (ricerca in griglia) - Convalida incrociata nidificata (ricerca in grigliacv) - Metriche delle prestazioni
- Sintesi

Laboratorio: Sperimentazione di pratiche di machine learning su casi concreti

5 - Apprendimento d'insieme

- Introduzione
- L'approccio al voto
- Una variante: impilabile
- Insaccamento
- Foreste casuali
- Potenziamento
- La variante Adaboost
- Aumento del gradiente
- Schede riassuntive

Laboratorio: Apprendimento complessivo su un caso concreto

6 - Regressione

- Regressione lineare semplice
- Regressione lineare multivariata
- Relazioni tra variabili
- Valori anomali (RANSAC)
- Valutazione delle prestazioni del modello di regressione
- Regolarizzazione dei modelli di regressione lineare
- Regressione polinomiale
- Regressione con foreste casuali
- Sintesi

Laboratorio: Regressione su un caso concreto

7 - Raggruppamento

- Introduzione
- Raggruppamento di oggetti per somiglianza con k-means
- k-significa: algoritmo
- L'inerzia di un cluster
- Variante k-significa++
- Raggruppamento sfocato
- Trovare il numero ottimale di cluster con il metodo Elbow
- Comprendere la qualità dei cluster con il metodo della silhouette
- Raggruppamento gerarchico
- Clustering mediante misurazione della densità DBSCAN
- Altri approcci di clustering
- Sintesi

Laboratorio: Clustering su un caso concreto



Obiettivi del corso

- Comprendere le differenze tra apprendimento automatico supervisionato, non supervisionato e meta-aprendimento
- Saper trasformare un grande volume di dati eterogenei in informazioni utili
- Padroneggia l'uso di algoritmi di autoapprendimento adattati a una soluzione di analisi
- Comprendere come sfruttare grandi volumi di dati testuali
- Essere in grado di applicare queste diverse tecniche ai progetti Big Data



Esercitazioni

- Una formazione molto **pratica**: il 70% del tempo della formazione è dedicato alla sua messa in pratica per una migliore assimilazione delle nozioni di base
- Il lavoro pratico viene svolto principalmente con R e Python
- Consulenti **esperti** condividono il loro know-how con i partecipanti



Date 2026



Ultimi posti Edizione garantita

dal 8 apr al 10 apr

dal 26 set al 28 set

dal 20 lug al 22 lug

dal 14 nov al 16 nov