Base Knowledge of Statistic

Artificial Intelligence Master Course

PhD Student Stefano Pio Zingaro

14 maggio 2018

Department of Computer Science and Engineering
Università degli studi di Bologna

Table of contents

Premessa

Di che cosa parleremo

- Deep learning
- Feature Extraction
- PCA
- Supervised Learning
- Unsupervised Learning

Di che cosa NON parleremo

- Augmented reality
- Big data
- Embodied agent
- Logic programming
- Multi-agent system
- Spatial-temporal reasoning

Formulazione dell'Ipotesi

Ipotesi

- Quando si effettua un esperimento vengono formulate delle ipotesi.
- Per ognuna delle ipotesi, viene fissata una confidenza.
- È possibile valutare quanto i risultati dell'esperimento sono in accordo con l'ipotesi grazie ai test di significatività, calcolando il p_{value}(*livello di significatività osservato*).

Esempio

Se la confidenza desiderata è 95% allora $\alpha = 1 - 0.95 = 0.05$, α viene detto *livello di significatività atteso*

Workflow

1. Ipotesi H (non dovuto al caso)
   VS
   Ipotesi H_0 (dovuto al caso)
2. Risultati Sperimentali $X = (x_1, x_2, x_3, ..., x_n)$
3. Probabilità dei risultati data l'ipotesi $P(X|H_0)$
4. Test di Significatività
   - $(p_{value} > \alpha) \rightarrow$ **Non Rigetto** H_0
   - $(p_{value} <= \alpha) \rightarrow$ **Rigetto** H_0

Attenzione

Se l'ipotesi H_0 viene respinta al livello di significatività 5%, allora abbiamo il 5% di probabilità di respingere un'ipotesi che era vera.

Fonte: `http://www.quadernodiepidemiologia.it/epi/assoc/t_stu.htm`

Raccolta dei Dati

Raccolta dei Dati

Esempio

Ipotesi H

"L'età di una persona influisce sulla sua altezza"

Domanda

Quali dati sarà più utile raccogliere?

1. Altezza
2. Età
3. Sesso

Analisi Statistica dei Dati

Perché l'Analisi Statistica dei Dati?

Esperimento

Gli esperimenti producono un flusso di DATI.

Interpretazione

I dati, interpretati, possono dare un significato all'esperimento. Spesso vengono ricercate:

- Differenze tra le misurazioni;
- Invarianti del sistema studiato.

Analisi

Diventa necessaria un'attenta Analisi Statistica.

Rappresentazione dei Dati: Gli Istogrammi

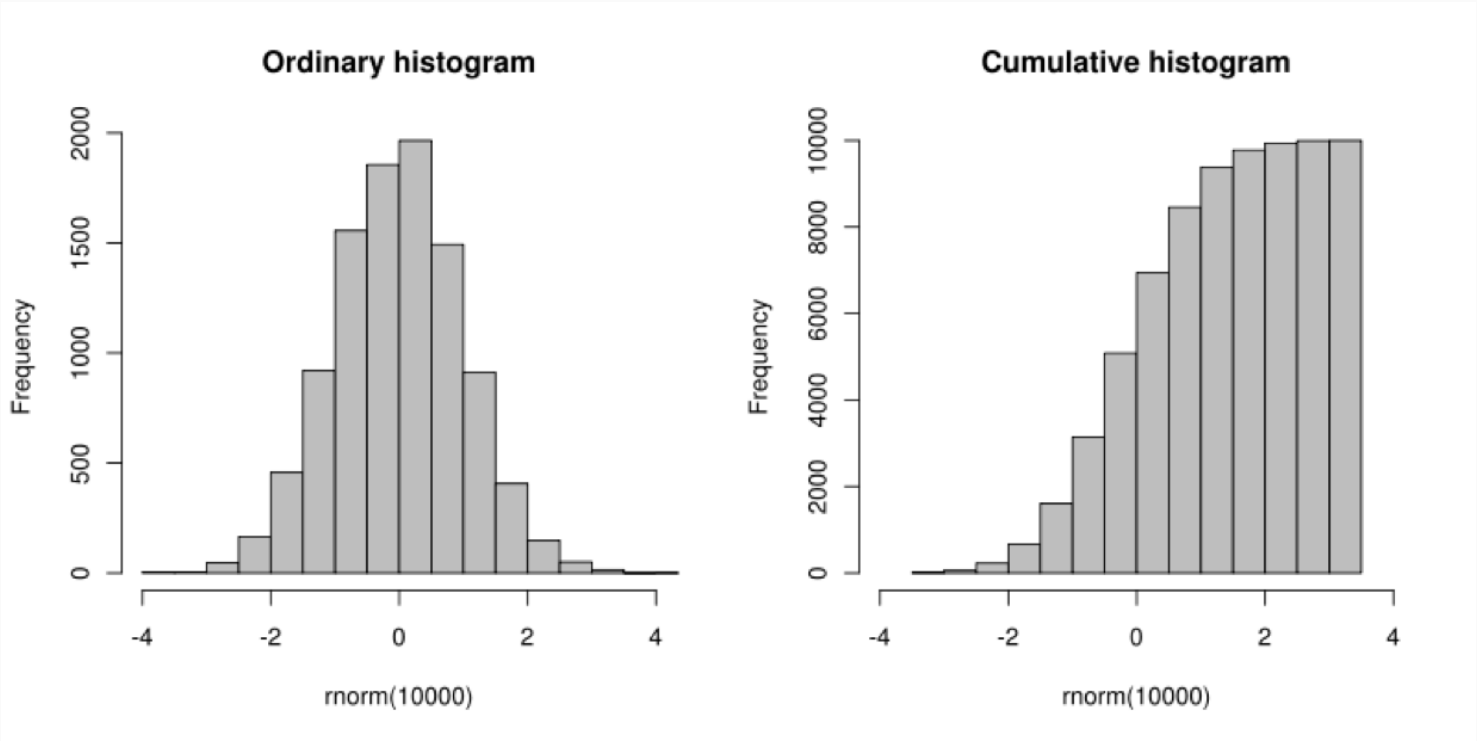

Figura 1: L'istogramma è una rappresentazione grafica di una distribuzione di frequenza di una certa grandezza, ossia di quante volte in un insieme di dati si ripete lo stesso valore.

Descrizione dei Dati: le Distribuzioni

- **Distribuzione Binomiale** Variabili che prendono i loro valori da un insieme discreto (eg. La probabilità di avere k eventi positivi su n eventi indipendenti).
- **Distribuzione Normale** Variabili che prendono i loro valori da un insieme continuo (eg. La probabilità di ottenere valori di x in un intervallo infinitesimo).

Distribuzione Binomiale per le Variabili Discrete

Definizione

$$P(k) = P(X_1 + X_2 + ... + X_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

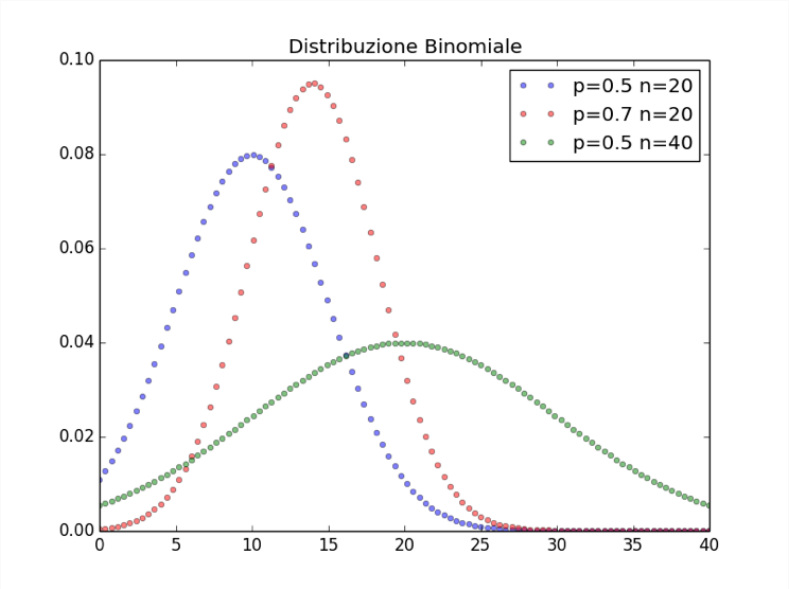

Figura 2: Distribuzioni binomiali con diversi parametri p e n.

Distribuzione Normale per le Variabili Continue

Definizione

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{con} \quad x \in \mathbb{R}.$$

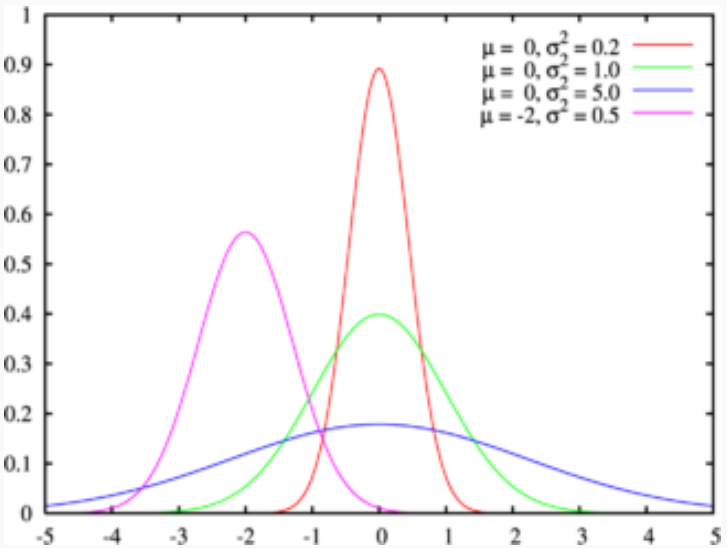

Figura 3: Gaussiana con vari parametri di media μ e varianza σ^2

Test di Significatività Parametrici

Definizione

Il test di significatività parametrico prevede il confronto tra due distribuzioni **normali** (eg. medie, varianze).

- Confronto delle Medie
  1. Z-test [1]
  2. T-Test [2]
- Confronto delle Varianze
  1. ANOVA [3]
  2. Test Chi-Quadrato [4]

[1] https://it.wikipedia.org/wiki/Test_Z
[2] https://it.wikipedia.org/wiki/Test_t
[3] https://it.wikipedia.org/wiki/Analisi_della_varianza
[4] https://it.wikipedia.org/wiki/Test_chi_quadrato

Test di Significatività NON Parametrici

Definizione

Il test di significatività NON parametrico prevede il confronto tra due distribuzioni **NON normali** (eg. medie, varianze).

- Test del Segno [5]
- Test di Wilcoxon [6]

[5] https://it.wikipedia.org/wiki/Test_dei_segni

[6] https://it.wikipedia.org/wiki/Test_dei_ranghi_con_segno_di_Wilcoxon

Estrazione delle Proprietà Rilevanti

- Correlazione di Pearson [7]
- Correlazione di Spearman [8]
- Principal Component Analysis (PCA) [9]

[7] https://it.wikipedia.org/wiki/Indice_di_correlazione_di_Pearson
[8] https://it.wikipedia.org/wiki/Coefficiente_di_correlazione_per_ranghi_di_Spearman
[9] https://it.wikipedia.org/wiki/Analisi_delle_componenti_principali

PCA: Principal Component Analysis

Definizione

L'analisi delle componenti principali è una tecnica per la semplificazione dei dati. Permette di scegliere in quante componenti **ridurre le dimensioni del problema**.

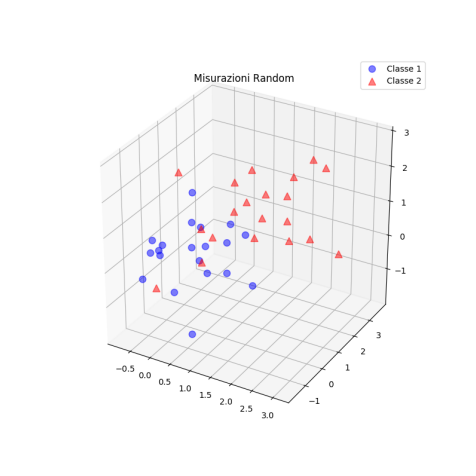

Figura 4: Rappresentazione grafica di due classi di punti a 3 dimensioni

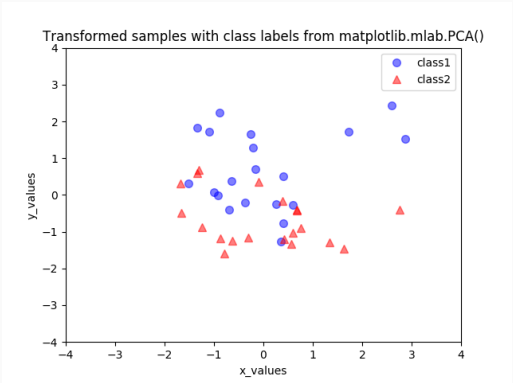

Figura 5: Rappresentazione grafica delle classi precedenti dopo l'applicazione di una PCA a 2 componenti.

Generalizzazione delle Ipotesi

Metodi di Generalizzazione delle Ipotesi

- Supervised (apprendimento da esempi reali)
  - Neural Networks [10]
  - Hidden Markov Models [11]
  - Support Vector Machines [12]
- Unsupervised (o semi-supervised/apprendimento ex-novo)
  - Hierarchical clustering [13]
  - K-means [14]

---

[10] https://it.wikipedia.org/wiki/Rete_neurale_artificiale
[11] https://it.wikipedia.org/wiki/Modello_di_Markov_nascosto
[12] https://it.wikipedia.org/wiki/Macchine_a_vettori_di_supporto
[13] https://it.wikipedia.org/wiki/Clustering_gerarchico
[14] https://it.wikipedia.org/wiki/K-means

Python's Machine Learning's Frameworks

Tensor Flow

- Libreria disponibile in diversi linguaggi di programmazione per il calcolo scientifico.
- Abilita la potenza della GPU (NVIDIA CUDA®) per il calcolo computazionale.
- Installazione in Python 2.7 con *pip*:
  1. `pip install tensorflow`
  2. `pip install tensorflow-gpu`
- Documentazione esaustiva e Tutorials disponibili su `https://www.tensorflow.org`.

Keras

- Installazione in Python 2.7 con *pip*:
  1. `pip install keras`
- Codice della libreria con esempi su `https://github.com/fchollet/keras/tree/master/keras`.

PyTorch

- Installazione in Python 2.7 con *pip*:
  1. `pip install http://download.pytorch.org/whl/torch-0.1.11.post5-cp27-none-macosx_10_7_x86_64.whl`
  2. `pip install torchvision`
- Documentazione ed esempi di codice su `http://pytorch.org/docs/`.

Conclusioni

Summary

Get the source of this work and the demo presentation from

`stefanopiozingaro.github.io/teaching.html`

Questions?