

Report del progetto di Machine Learning

Ferrario Tommaso Matr. 869005 (@TommasoFerrario18)

Terzi Telemaco Matr. 865981(@Tezze2001)

Vendramini Simone Matr. 866229(@Svendra4MySelf)

19 dicembre 2023

Indice

1	Introduzione	2
2	Dataset	3

Capitolo 1

Introduzione

Questo è un progetto per l'esame di Machine Learning del primo anno del corso di laurea magistrale in informatica dell'Università degli Studi di Milano-Bicocca.

L'intero progetto si basa sul riconoscimento della presenza di un tumore al cervello data l'immagine di una risonanza magnetica. Il dataset scelto per questo progetto è scaricabile dal seguente link.

Per il riconoscimento del tumore sono stati allenati i seguenti modelli di machine learning:

- **Percettrone:** è stato scelto questo modello vista la sua semplicità e il suo ridotto numero di parametri rispetto agli altri.
- **SVM:** è stato scelto questo modello vista la buona capacità teorica nel generalizzare.
- **Naive Bayes Gaussiano:** è stato scelto questo modello dal momento che è l'unico ad essere probabilistico.

La relazione è stata suddivisa nei seguenti capitoli:

- **Introduzione:** descrizione del dominio e presentazione dei modelli che verranno presi in considerazione per questo progetto.
- **Dataset:** descrizione di come è stato costruito il dataset a partire dalle immagini, ovvero come sono state ricavate le features, e analisi esplorativa.
- **Percettrone:** descrizione e analisi delle performance del percertrone.
- **SVM:** descrizione e analisi delle performance delle SVM.
- **Naive Bayes Gaussiano:** descrizione e analisi delle performance per Naive Bayes Gaussian.
- **Analisi dei risultati:** analisi comparata dei risultati tra i tre modelli considerati.
- **Conclusioni:** conclusioni sull'elaborato.

Capitolo 2

Dataset

Il dataset è stato a partire da un set di 3762 immagini ottenute dalla risonanza magnetica del cervello di 3762 persone, etichettato manualmente da professionisti del settore nelle rispettive classi:

- **presenza del tumore:** $T = 1$
- **assenza del tumore:** $T = 0$

Il valore della label cade sotto al colonna *Class*.

Le features del dataset sono state ottenute calcolando i **momenti Hu** sulle immagini della risonanza magnetica. I momenti Hu catturano le informazioni di base sull'immagine come l'area dell'oggetto, il centroide, l'orientazione e altre proprietà.

Le feature sulle immagini si dividono in base a 2 gruppi[1]:

- **First Order Features:** forniscono informazioni legate alla distribuzione dei livelli di grigio dell'immagine. Queste features corrispondono alle statistiche descrittive calcolate sui valori di ciascun pixel dell'immagine:
 - **media**
 - **varianza**
 - **deviazione standard**
 - **indice di asimmetria**
 - **indice di kurtosis**
- **Second Order Features:** forniscono informazioni a livello di composizione della texture dell'immagine.
 - **contrast**
 - **energy**
 - **asm**
 - **entropy**
 - **homogeneous**
 - **dissimilarity**
 - **correlation**
 - **coarseness**

Bibliografia

- [1] Namita Aggarwal e RK Agrawal. “First and second order statistics features for classification of magnetic resonance brain images”. In: (2012).