



Università degli Studi dell'Aquila
Dipartimento di Ingegneria e Scienze dell'Informazione e Matematica

Corso di Machine Learning

Homework 3: Manuale

Angelo Damiani

Anno Accademico 2017/2018

Indice

1	Installazione	3
2	Funzioni	4
2.1	Procedura di allenamento e testing	4
2.2	Procedura di testing	6
2.3	Procedura di interrogazione	6

Capitolo 1

Installazione

Per l'avvio dell'applicazione è necessaria l'installazione di Python v3.6 e delle seguenti librerie:

- numpy;
- tensorflow;
- keras;
- h5py (per la lettura e la scrittura dei pesi della rete neurale).

Per l'installazione di queste librerie è necessario il programma `pip3.6`, il pacchetto manager di Python v3.6. Eseguire il seguente codice per installare `pip3.6` su piattaforma Linux :

```
sudo add-apt-repository ppa:jonathonf/python-3.6
sudo apt update
sudo apt install python3.6
sudo apt install python3.6-dev
wget https://bootstrap.pypa.io/get-pip.py
sudo python3.6 get-pip.py
```

Per installare le librerie eseguire:

```
sudo pip3.6 install tensorflow==1.3.0
sudo pip3.6 install keras
sudo pip3.6 install h5py
```

È necessario che tensorflow sia alla versione 1.3.0 poichè la versione 1.4.0 non supporta Python 3.6 (con il quale è stato sviluppato l'homework).

Capitolo 2

Funzioni

A dispetto da ciò che è stato richiesto per l’homework, sono state sviluppate 3 diverse funzioni all’interno del main: allenamento di una rete neurale (con valutazione finale delle performance, ciò che è stato richiesto), test delle performance di una rete neurale preesistente, query di una rete neurale preesistente. Per il resto della descrizione delle funzionalità si simulerà l’esecuzione con il dataset UCI Iris.

Quindi, per avviare l’applicazione andare via terminale nella cartella e avviare `main.py` tramite:

```
python3.6 main.py
```

Dopo un breve caricamento della libreria si presenterà la scritta:

“What do you want to do? (‘Q’ to query an already existing neural network, ‘T’ to train a new one, ‘V’ to test an already existing neural network, anything else to close.”

Ovviamente premendo “Q” si avvierà la procedura per la query, premendo “T” si avvierà la procedura per l’allenamento di una rete neurale e premendo “V” si avvierà la procedura per la valutazione.

2.1 Procedura di allenamento e testing

Premendo ‘T’, dunque, si avvia la procedura di allenamento della rete neurale.

La prima cosa che viene richiesta è il nome che si vuol dare alla rete. Il nome inserito verrà utilizzato per salvare: la rete neurale, i campioni da utilizzare per il test set (per usi successivi) e la discretizzazione.

Dopo aver inserito il nome da dare (per esempio **iris**) verrà richiesto di immettere il path per il dataset con il quale costituire la rete. Nel caso in esame, essendo il dataset all'interno della cartella, si digita: `files/UCI Repository/iris.data`.

Ora viene richiesto di inserire il numero di feature obiettivo. Da descrizione dell'homework ogni feature è feature obiettivo e dunque si dovrà inserire il numero di feature del generico campione. Ad esempio per il dataset UCI Iris tale numero è **5**.

Una volta inserito il numero verranno richieste una ad una le feature obiettivo richieste e dunque si dovrà inserire: **0, 1, 2, 3, 4**.

A questo punto verrà richiesto il numero di layer che si desiderano per la rete neurale. Si ipotizzi di volerne **5** (al di fuori di quello di input, che come detto nel report, è sottinteso e quello di output che è calcolato automaticamente). Dopo aver inserito il numero di layer desiderato viene richiesto, per ciascuno di questi layer; ipotizziamo: **100, 200, 300, 200, 100**. Adesso viene richiesto il numero di campioni con cui allenare la rete neurale; si supponga di volerne **30000**. Infine, le ultime due cose richieste, sono il numero di epoche con il quale allenare la rete (**100**) e il numero di campioni con il quale testare la rete (**2000**).

In questo momento la rete comincia ad allenarsi e, alla fine dell'allenamento, viene testata. Alla fine del testing vengono mostrate le statistiche. Il risultato rispetto ai dati inseriti è dato dalla figura 2.1.

```
Neural network trained
Test performed
Exact predictions = 630
Exact predictions percentage = 31.5%
Mean distance = 1.6100527159320164
Mean error on single feature = 0.32201054318640326
```

Figura 2.1: Risultato con i dati d'esempio.

Alla fine della procedura, avendo selezionato il nome **iris**, verranno creati i seguenti file (con le relative cartelle):

- `files/discretizations/iris/dscr.pkl` e `files/discretizations/iris/type.pkl`: contengono la discretizzazione del dataset UCI Iris;

- `files/NeuralNetworks/iris/model.json` e `files/NeuralNetworks/iris/weights.h5`: contengono i parametri e i pesi della rete neurale; necessari per caricare la rete.
- `files/testsets/iris/testset`: contenente i campioni del dataset che non sono stati usati per il training. È possibile utilizzare questo campione come test set.

2.2 Procedura di testing

Premendo ‘V’, invece, si avvia la procedura di testing di una rete neurale preesistente. La prima cosa che viene richiesta, ancora, è il nome della rete (**iris**). Una volta inserito il nome, viene richiesto il path per il testset da utilizzare. Si ipotizzi di voler usare il testset generato dalla procedura di training. Deve essere inserito, allora, `files/testsets/iris/testset`.

A questo punto viene richiesto il numero di campioni su cui testare la rete, supponiamo **2000**.

Fatto ciò viene richiesto il numero di variabili obiettivo e, come nel training, deve essere reinserito **5**. Una volta inserito viene chiesto di esplicitare quali sono queste variabili: **0, 1, 2, 3, 4**.

Infine verrà testata la rete neurale e verranno mostrate le statistiche. Tali statistiche riguardano la distanza media tra l’output e ciò che ci si aspettava e il numero di matching perfetti avvenuti.

2.3 Procedura di interrogazione

Premendo ‘Q’, invece, si avvia la procedura di interrogazione di una rete preesistente. Inizialmente viene richiesto il nome della rete da interrogare, di nuovo inseriamo “**iris**”.

A questo punto viene richiesto il path su cui interrogare la rete neurale. Il file in questione dovrà essere nello stesso formato del dataset ma con dei punti interrogativi (“?”) in alcune posizioni. Insieme al software è allegato un file di query (per iris) semplificato, estratto dal testset: `files/queryiris`.

Una volta inserito verrà salvato l'output nella cartella `files/output/iris/output`.