22. Overfitting e regolarizzazione

Corso di Python per il Calcolo Scientifico

Angelo Cardellicchio
angelo.cardellicchio@stiima.cnr.it

Outline

- Overfitting ed underfitting
- Regolarizzazione
- Layer di dropout

Overfitting ed underfitting

- Normalmente, oltre una certa epoca di training, l'accuracy sui dati di validazione **tende a diminuire**.
- Questo è legato al fatto che il modello aderisce *eccessivamente* ai dati di training, e comporta il fenomeno dell'**overfitting**.
- Contrapposto all'overfitting è l'**underfitting**, legato a situazioni nelle quali la rete non raggiunge la massima accuracy possibile.
- Mitigare l'underfitting significa usare più parametri o un maggior numero di epoche di training; mitigare l'overfitting significa usare un dataset più ampio o delle tecniche di **regolarizzazione**.

Regolarizzazione

- La regolarizzazione introduce un componente nella funzione di costo che penalizza pesi eccessivi, rendendo i diversi neuroni più *uniformi* nella loro risposta.
- L'uso è giustificato dal **rasoio di Occam**: dati due modelli che spiegano un fenomeno, quello più semplice sarà anche quello più efficace.
- In TensorFlow e Keras, ciò avviene usando il parametro **`kernel_regularizers`** ed un oggetto di classe **`l2`** o **`l1`**.

```
from keras import regularizers

layers.Dense(
    64,
    activation='relu',
    kernel_regularizers=regularizers.l2(0.001))
```

Layer di dropout

- Il layer di **dropout** permette di isolare il contributo effettivo di ciascun nodo della rete.
- Per farlo, si *disconnettono*, in maniera casuale, un certo numero di neuroni in ingresso dallo strato precedente ad ogni epoca di training.
- Modificando la connettività di ciascun neurone, viene introdotto del rumore nella fase di apprendimento, per cui si riesce ad evitare che i neuroni si configurino per evitare errori propri dei dati sotto esame.
- Normalmente, la percentuale dei neuroni del layer precedente ignorati ad ogni epoca di training è del 50%.
- Keras ha un layer apposito per il dropout.

Domande?

42