

Machine Learning

29 Aprile, 2024

Nome:

Numero di matricola:

Il punteggio massimo totale è 35.

Domanda 1 (3 punti) Immagina di aver addestrato due classificatori usando la cross-entropy loss:

- In un caso hai misurato solo l'*accuracy*, che a convergenza arriva al 100%. Questo implica che la loss sia a sua volta uguale a zero?
- Nel secondo caso hai misurato solo la *loss* e osservi che raggiunge lo zero numerico. Questo implica che l'*accuracy* sia al 100%?

Motivare le risposte.

Domanda 2 (3 punti) Usando *broadcasting*, scrivi la seguente funzione: data in input una matrice di n punti $\mathbf{X} \in \mathbb{R}^{n \times k}$, restituisci in output una matrice \mathbf{D} di dimensioni $n \times n$ contenente le distanze tra tutte le coppie di punti.

Domanda 3 (6 punti) Hai un dataset di 100 punti distribuiti casualmente in modo uniforme in un'ipersfera in 10,000 dimensioni. Questi punti sono codificati in una matrice \mathbf{X} .

- Ti aspetti di osservare una struttura particolare in $\mathbf{X}^\top \mathbf{X}$?
- Immagina di voler fare PCA, quindi consideri il prodotto $\mathbf{X}\mathbf{X}^\top$. Ti aspetti un comportamento particolare dalle componenti principali?

Domanda 4 (3 punti) Considera un MLP con bias, che trasforma un input 100-dimensionale in questo modo: $100 \rightarrow 256 \rightarrow 512 \rightarrow 10$. Quanti pesi ha in tutto?

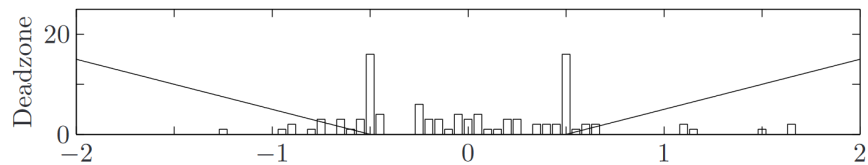
Domanda 5 (4 punti) Da cosa dipende il “landscape energetico” di una rete neurale?

- ☐ Il dataset
- ☐ Il numero di layers
- ☐ La funzione di loss
- ☐ I parametri della rete
- ☐ Le funzioni di attivazione
- ☐ Tutte le precedenti

Motivare le risposte.

Domanda 6 (6 punti) Consideriamo una variante dell'algoritmo di embedding SNE, dove la matrice di probabilità \mathbf{P} viene modificata sostituendo $p_{ij} = \frac{p_{ij} + p_{ji}}{2n}$. Come cambia il gradiente di SNE alla luce di questa modifica?

Domanda 7 (4 punti) La funzione di penalty chiamata "deadzone" ha un comportamento di questo tipo:



- Descrivere a parole l'illustrazione, spiegando nel dettaglio in che modo agisce la penalty e che tipo di regolarizzazione applica all'input.
- Scrivere un'espressione matematica della deadzone, basandosi unicamente sul grafico.

Domanda 8 (6 punti) Un membro del vostro gruppo di studio ha avuto un'idea per un nuovo algoritmo di training da usare al posto di SGD – ma non ha tempo per spiegare, e vi invia solamente il codice dicendo al volo “questo va più veloce!”. Ecco il codice:

```
1  grads_squared = 0
2  for _ in num_iterations:
3      dw = compute_gradient(x, y)
4      grad_squared += dw * dw
5      w = w - (lr / np.sqrt(grad_squared)) * dw
```

Qual è l'idea dietro questo codice? Descrivi cosa fa e spiega perché dovrebbe portare ad un training più efficiente. Se pensi non possa funzionare in alcuni casi, spiega perché. In generale, fai del tuo meglio per dimostrare di aver capito.