UNIVERSITÀ DEGLI STUDI DI PALERMO

DIPARTIMENTO DI INGEGNERIA

LAUREA MAGISTRALE IN INGEGNERIA INFORMATICA - INTELLIGENZA ARTIFICIALE

TECNICHE DI AML PER LA GENERAZIONE DI ATTACCHI QUERY EFFICIENT

Tesi di Laurea di
Giuseppe Ganci

Relatore:
Prof. Marco Morana

Controrelatore:
Prof. ?

Correlatore:
?

Anno Accademico 2022/2023

Indice

Introduzione

Capitolo 1

Stato dell'arte

Capitolo 2

Formulazione matematica del problema

Un problema di Neural Architecture Search (NAS) consiste in una ricerca in uno spazio degli stati, a partire da una architettura di partenza, detta **Supernet**, per identificare, tra tutte le possibili **Subnet**, quella che massimizza una determinata funzione obiettivo. In questo caso, al fine di ottenere una compressione di modelli *Transformer*, la ricerca si configura come un problema di **ottimizzazione multi-obiettivo**, tramite il quale vengono ottimizzate sia la performance del modello (attraverso l'accuratezza) sia la dimensione (attraverso il numero di parametri).
Uno dei problemi principali nell'ambito NAS è proprio la dimensione dello spazio degli stati, che risulta elevatissima, poiché il numero di possibili configurazioni cresce in maniera combinatoria rispetto al numero di componenti del modello. Nei Vision Transformer, ogni blocco Transformer presenta molteplici gradi di libertà: è possibile variare il numero di teste di attenzione nell'unità di *Multi-Head Attention* (MHSA), la dimensione dei vettori *Query-Key* e *Value*, ma anche la dimensione del livello nascosto nel modulo *Multi-Layer Perceptron* (MLP), ecc... Anche considerando un numero limitato di opzioni per ogni strato, il numero totale di subnet candidate diventa rapidamente intrattabile per una ricerca esaustiva.
Proprio per rendere trattabile l'esplorazione dello spazio degli stati è stata scelta una formulazione conveniente del problema, attraverso un algoritmo di ricerca *Branch and Bound* e l'utilizzo di azioni di *pruning* predefinite. Di seguito la formulazione matematica nel dettaglio.

2.1 Formulazione della ricerca

Dato un modello ViT M con parametri θ, definiamo un processo iterativo di pruning. Per ogni iterazione viene eseguita una ricerca di tipo *Depth First* a profondità limitata, ottimizzata tramite algoritmo *Branch and Bound*. Durante la ricerca verranno selezionate delle azioni da eseguire sul modello, e successivamente, attraverso un secondo problema combinatorio vengono selezionate le specifiche dimensioni da eliminare. L'obiettivo è massimizzare la seguente funzione rispetto alla maschera binaria m, ai parametri θ e al dataset di validazione D:

$$\max_{m} \text{Obj}(m, \theta, D) = \log_2(\mathscr{A}(m, \theta, D)) - \lambda \cdot \log_2 \left(\frac{\mathscr{P}(m, \theta)}{\mathscr{P}_{tot}(\theta)} \right) \tag{2.1}$$

Dove:

- $m \in \{0,1\}^N$ è la maschera di pruning globale, con N numero totale di parametri.
- La maschera è strutturata per blocchi: $m = \{m_{\text{emb}}, m^{(1)}, \ldots, m^{(i)}, \ldots, m^{(L)}\}$, con L numero di layer.
- Per ogni blocco i: $m^{(i)} = \{m_{\text{head}}, m_{\text{attn}}, m_{\text{v\_proj}}, m_{\text{mlp}}\}$.
- $\mathscr{A}(m, \theta, D)$ indica l'accuratezza del modello mascherato sul dataset D.
- $\mathscr{P}(m, \theta)$ rappresenta il numero di parametri attivi (non zero).
- $\mathscr{P}_{tot}(\theta)$ rappresenta il numero totale di parametri del modello originale.

Notare che tramite questa formulazione è possibile ricondurre il problema in esame ad un **Problema di Programmazione Non Lineare Binario** (PNLPB). In questo contesto, le variabili decisionali sono rappresentate dai componenti della maschera $m \in \{0,1\}^N$, mentre la non linearità è introdotta sia dalla natura della funzione di accuratezza $\mathscr{A}$ (legata ai pesi della rete neurale) sia dall'operatore logaritmico utilizzato nella funzione obiettivo. Nonostante esistano in letteratura delle tecniche di *Smoothing* delle variabili binarie per risolvere questo tipo di problemi combinatori, come quelle utilizzate da Murray et al. [2], i tempi di risoluzione, con un numero ridotto di variabili binarie rispetto al caso in esame, sono molto elevati. Proprio per tale motivo si è scelto di utilizzare un differente approccio algoritmico basato sul *Branch and Bound*.

2.1.1 Branching

L'esplorazione dello spazio di ricerca avviene tramite la costruzione di un apposito albero, seguendo un algoritmo di esplorazione *Depth First* a profondità limitata; di conseguenza, è possibile definire la strategia di esplorazione come **Depth Limited Depth First Branch and Bound** (DL-DFBnB).
La radice di questo albero è rappresentata dalla *Supernet*, ovvero il modello *Vision Transformer* di dimensione originale, specializzato su uno specifico dominio tramite *fine-tuning*. A partire dalla radice, viene eseguita la fase di **Branching** applicando le azioni di seguito elencate:

- **Pruning Multi-Layer Perceptron**
- **Pruning Query-Key**
- **Pruning Value-Projection**
- **Pruning Head**
- **Pruning Embedding**

Si noti che tutte le azioni, ad eccezione del *Pruning* dell'*Embedding*, sono locali a uno specifico blocco *Transformer*. La formalizzazione matematica delle singole azioni è rimandata alla sezione successiva.
Attraverso il *branching* viene costruito un **albero 5-ario** di ricerca, dove ogni nodo rappresenta una specifica *Subnet* ottenuta a partire dal nodo genitore. Ad ogni nodo viene associato un valore della funzione obiettivo, calcolato valutando l'accuratezza sul *Search Set* (utilizzando esclusivamente i parametri attivi) e il relativo numero di parametri residui. Questo valore verrà successivamente utilizzato per verificare i vincoli di *Bound* e, qualora risultasse più elevato del miglior valore individuato dalla ricerca, si aggiornerebbe quest'ultimo e si salverebbe la maschera di *Pruning* corrispondente.
Al fine di massimizzare l'efficacia della potatura, la strategia di esplorazione *Depth First* è stata scelta per favorire il raggiungimento dei nodi foglia dell'albero di ricerca. Questo permette all'algoritmo di individuare rapidamente configurazioni caratterizzate da un elevato numero di azioni di *pruning* e, di conseguenza, da una significativa riduzione dei parametri.
L'introduzione di un limite di profondità nell'albero di ricerca permette di mitigare l'impatto cumulativo del *pruning* sulle prestazioni del modello. Un'esplorazione eccessivamente profonda,

infatti, comporterebbe una rimozione massiva di parametri, rischiando di degradare l'accuratezza in modo irreversibile. In tali scenari, il danno strutturale all'architettura potrebbe risultare troppo elevato, rendendo difficoltoso il recupero delle performance originali anche attraverso l'impiego di tecniche avanzate come la *Knowledge Distillation*.

2.1.2 Bound

2.1.3 Formulazione delle Azioni di Pruning

La selezione dei parametri da rimuovere è modellata come un problema di minimizzazione locale volto a identificare il gruppo di parametri g la cui rimozione minimizza l'impatto sulla funzione di perdita $\mathscr{L}$. Utilizziamo un'approssimazione di Taylor del primo ordine per stimare la variazione della loss, elevata al quadrato per considerare la magnitudine dell'impatto. Definiamo la metrica di importanza $\mathscr{I}(g)$ per un gruppo g come:

$$\mathscr{I}(g) = \left(\sum_{w \in g} w \cdot \frac{\partial \mathscr{L}}{\partial w} \right)^2 \tag{2.2}$$

dove il termine $\sum w \cdot \frac{\partial \mathscr{L}}{\partial w}$ rappresenta il prodotto scalare tra i pesi e i loro gradienti.

Per garantire l'efficienza hardware (es. Tensor Core), imponiamo un vincolo di cardinalità $|g| = K$ (con $K = 8$ o $K = 32$). Le azioni di ricerca sono formalizzate come problemi di minimizzazione sull'insieme $\mathscr{S}_K(\mathscr{D}) = \{g \subset \mathscr{D} : |g| = K\}$, che rappresenta la famiglia di tutti i possibili sottoinsiemi di parametri nello spazio $\mathscr{D}$ con cardinalità K:

- **Pruning MLP:** Selezione dei $K = 32$ neuroni meno rilevanti nello spazio dei neuroni attivi $\mathscr{D}_{\text{MLP}}$ del blocco corrente:

$$g^*_{\text{MLP}} = \arg \min_{g \in \mathscr{S}_{32}(\mathscr{D}_{\text{MLP}})} \mathscr{I}(g) \tag{2.3}$$

- **Pruning QK (Query-Key):** Selezione delle $K = 8$ dimensioni meno rilevanti nello spazio delle feature delle teste di attenzione $\mathscr{D}_{\text{QK}}$:

$$g^*_{\text{QK}} = \arg \min_{g \in \mathscr{S}_8(\mathscr{D}_{\text{QK}})} \mathscr{I}(g) \tag{2.4}$$

- **Pruning VPROJ (Value-Projection):** Selezione delle $K = 8$ dimensioni meno rilevanti nello spazio di proiezione dei valori $\mathscr{D}_{\mathrm{V}}$:

$$g^*_{\mathrm{vproj}} = \arg \min_{g \in \mathscr{S}_8(\mathscr{D}_{\mathrm{V}})} \mathscr{I}(g) \tag{2.5}$$

- **Pruning HEAD:** Rimozione di un'intera testa di attenzione h, dove il gruppo g_h comprende tutti i pesi associati a quella testa:

$$h^* = \arg \min_{h \in \mathrm{Heads}} \mathscr{I}(g_h) \tag{2.6}$$

- **Pruning EMB (Embedding):** Selezione delle $K = 8$ dimensioni meno rilevanti nello spazio dell'embedding globale $\mathscr{D}_{\mathrm{emb}}$, valutando il contributo aggregato su tutti i layer:

$$g^*_{\mathrm{emb}} = \arg \min_{g \in \mathscr{S}_8(\mathscr{D}_{\mathrm{emb}})} \mathscr{I}(g) \tag{2.7}$$

Capitolo 3

Metodo adottato

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

3.1 Logica comune ai sistemi di analisi

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur non nunc in purus aliquam eleifend. Sed sed justo in nisl fringilla vehicula. Aenean sodales pellentesque porttitor. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Morbi lacus lacus, laoreet quis pretium non, dignissim sit amet purus. Vivamus eu mauris felis, eget vulputate metus. Vestibulum sollicitudin nisi vitae quam venenatis id ornare magna condimentum. Nullam faucibus commodo dui quis tempor. Donec sed placerat odio. Sed varius, mi et volutpat egestas, leo arcu laoreet ante, in commodo ipsum neque eget ligula. Fusce cursus justo at sem auctor dapibus. Proin eget quam sed orci eleifend molestie. Suspendisse malesuada velit eget lectus porttitor iaculis.

3.1.1 Selezione delle features

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco

laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

3.2 Tecniche di manipolazione dei sorgenti

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur non nunc in purus aliquam eleifend. Sed sed justo in nisl fringilla vehicula. Aenean sodales pellentesque porttitor. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Morbi lacus lacus, laoreet quis pretium non, dignissim sit amet purus. Vivamus eu mauris felis, eget vulputate metus. Vestibulum sollicitudin nisi vitae quam venenatis id ornare magna condimentum. Nullam faucibus commodo dui quis tempor. Donec sed placerat odio. Sed varius, mi et volutpat egestas, leo arcu laoreet ante, in commodo ipsum neque eget ligula. Fusce cursus justo at sem auctor dapibus. Proin eget quam sed orci eleifend molestie. Suspendisse malesuada velit eget lectus porttitor iaculis.

3.2.1 Manipolazioni singole

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

3.2.2 Manipolazioni Multiple

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Capitolo 4

Esperimenti e risultati

In questo capitolo vengono presentati i risultati ottenuti dal framework realizzato, tramite una serie di esperimenti che coinvolgono modelli e dataset differenti.
Tutti gli esperimenti vengono svolti con modelli **ViT**, preaddestrati sul dataset **ImageNet**, che lavorano con *patch* di dimensione **16×16** e immagini in *input* di tipo **RGB** con risoluzione **224×224**. La procedura utilizzata consta di due fasi:

- **FineTuning**: il modello preaddestrato su *ImageNet* viene sottoposto ad una fase di *Fine-Tuning* sul dataset di riferimento per l'esperimento, in questo modo si ottiene allo stesso tempo una **baseline** per confrontare le performance, ma anche un modello di partenza sul quale eseguire il framework di compressione.
- **Compressione**: una volta ottenuto il nuovo modello specializzato sul dominio del problema, viene eseguito il framework di **NAS iterativo** sviluppato, e vengono infine analizzate le performance sul *set* di validazione del dataset di riferimento.

Di seguito vengono presentati i risultati dei vari esperimenti effettuati.

4.1 Test sul dataset CIFAR-100

Il primo dataset utilizzato per testare il framework è stato il **CIFAR100**, che è stato selezionato poiché molto diffuso in letteratura, in cui è utilizzato come **benchmark standard** per compiti di classificazione. Risulta più complesso rispetto al *CIFAR10*, ma mantiene dimensioni ridotte e quindi permette una ottima rapidità di *training*. In particolare *CIFAR100* contiene **60.000**

immagini totali, suddivise in **50.000** immagini di addestramento e **10.000** di validazione, tutte della risoluzione di **32×32** e di tipo **RGB**.
Data la ridotta dimensione del dataset, e la grande capacità dei modelli *Vision Transformer*, per scongiurare il rischio di *overfitting*, sono state adottate delle tecniche di **Data Augmentation**, in particolare:

- **Random Resized Crop**: Questa tecnica estrae una porzione casuale dell'immagine originale con un'area compresa tra l'**80%** e il **100%** della dimensione iniziale. Successivamente, il ritaglio viene ridimensionato alla risoluzione *target* di 224×224 pixel.
- **Random Horizontal Flip**: Consiste nel riflettere l'immagine orizzontalmente con una probabilità del **50%**.
- **Color Jitter**: Questa trasformazione applica variazioni casuali alla luminosità, al contrasto e alla saturazione dell'immagine, con un fattore di distorsione impostato a **0.3** per ciascun parametro.

Per i test sul dataset *CIFAR100* è stata effettuata una ulteriore suddivisione del *set* di addestramento con rapporto **90/10**, in modo tale da ottenere un **Train Set** finale di **45.000** immagini e un **Held-Out Set** di **5.000** immagini, quest'ultimo utilizzato per il monitoraggio del processo di compressione.
Al fine di rendere il processo di *Neural Architecture Search* (**NAS**) efficiente e rendere ragionevoli i tempi di esecuzione, ad ogni iterazione viene campionato dal *Train Set* un insieme di **25 immagini per classe** (**2500** totali), in modo da ottenere una buona stima del gradiente, senza compromettere la velocità di esplorazione dello spazio architetturale. Questo insieme costituisce il **Search Set** ed è soggetto alle stesse tecniche di *Augmentation* viste in precedenza, in modo da evitare che il framework individui architetture carenti in termini di **robustezza** e **generalizzazione**. Il campionamento dinamico del *Search Set* ad ogni iterazione è una scelta progettuale essenziale per garantire la robustezza del processo di ricerca. Questa strategia evita che l'algoritmo converga verso architetture eccessivamente specializzate su un *set* statico di campioni, promuovendo invece l'individuazione di *subnet* capaci di generalizzare correttamente sull'intero dominio di *CIFAR-100*. Il ruolo di questo *Search Set* è duplice:

- **Stima della metrica di Importanza**: viene utilizzato per calcolare il valore di importanza di ogni parametro di una *subnet*, corrispondente ad uno specifico nodo della ricerca *Branch and Bound*.

- **Valutazione della Funzione Obiettivo**: viene utilizzato per calcolare l'accuratezza della *subnet*, coinvolta nel calcolo della funzione obiettivo del processo di ricerca.

4.1.1 Test con ViT-Small

In questi test è stato utilizzato un modello **ViT-Small** da **21.7 Mln** di parametri, a cui è stata aggiunta una testa di classificazione da **100 unità**, e successivamente sottoposto a *Fine Tuning* iniziale sul dataset in oggetto, con i seguenti iperparametri:

- **Ottimizzatore AdamW** con un fattore di **decadimento dei pesi** pari a **0.05** e con **learning rate discriminativi**: 0.5×10^{-5} per il *backbone* del *Transformer* e 0.5×10^{-4} per la testa di classificazione.
- **Scheduler** dei *learning rate* di tipo **Cosine Annealing**.
- Dimensione dei **batch** di *training* pari a **128** campioni.
- **30 epoche** di addestramento con **early stopping**.

Per quanto riguarda il framework, tutti i test sono stati effettuati con **15 iterazioni** di compressione. In ogni iterazione, la **profondità** dell'albero di ricerca è stata limitata a **6**, questo implica che, per ogni iterazione, al modello possono essere applicate massimo **6 azioni consecutive** di *pruning*. Viene inoltre utilizzata una **soglia di tolleranza** della fase di *Bound* pari a **0.005**, in modo tale da permettere una ricerca più approfondita dello spazio di ricerca, al costo di un numero lievemente maggiore di nodi esplorati. Infine, la **funzione obiettivo** utilizza un valore di λ pari a **1.0**, attribuito al contributo dei parametri.
La fase di **Recovery Fine-Tuning**, volta a ripristinare le prestazioni dopo il taglio dei parametri, adotta la medesima configurazione di ottimizzazione del *Fine-Tuning* iniziale, con due variazioni: l'impiego di un **learning rate unico** pari a 0.5×10^{-5} per l'intera rete e un limite massimo di **20 epoche** di addestramento, sufficienti per stabilizzare i pesi della *subnet* individuata.
Infine, per quanto concerne i test con la **Knowledge Distillation** (**KD**), è stata implementata una versione *logit-based* seguendo l'approccio proposto da *Hinton et al.* [1]. In questo caso si è scelto di utilizzare il *ViT-Small baseline* come modello **teacher**. La funzione di **loss composita** utilizzata durante il *Recovery Fine-Tuning* integra un termine di distillazione con temperatura $\tau = 4.0$. A tale componente è stato assegnato un **peso relativo pari a** 0.9, privilegiando così il

trasferimento della conoscenza dal modello *teacher* alla *subnet* compressa. Questa configurazione ha permesso di ammorbidire le distribuzioni di probabilità dei *logits*, consentendo al modello *student* di apprendere non solo le etichette corrette, ma anche le relazioni strutturali tra le classi catturate dal modello completo. Di seguito i risultati dei test.

Tabella 4.1: Risultati della compressione di ViT-Small su CIFAR-100.

Modello	**Parametri (M)**	**Top-1 Acc. (%)**	**GFLOPs**	**Throughput (img/s)**	**Latency (ms)**
Baseline	21.67	90.42%	9.20	1137.5	112.5
Pruned	**15.93** (-26.6%)	**89.59%** (-0.83%)	**6.85** (-25.51%)	**1405.5** (+23.56%)	**91.1** (-19.07%)

Capitolo 5

Caso di studio: deploy su mobile

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

5.1 Raze the ground a confronto con MCTS e BS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur non nunc in purus aliquam eleifend. Sed sed justo in nisl fringilla vehicula. Aenean sodales pellentesque porttitor. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Morbi lacus lacus, laoreet quis pretium non, dignissim sit amet purus. Vivamus eu mauris felis, eget vulputate metus. Vestibulum sollicitudin nisi vitae quam venenatis id ornare magna condimentum. Nullam faucibus commodo dui quis tempor. Donec sed placerat odio. Sed varius, mi et volutpat egestas, leo arcu laoreet ante, in commodo ipsum neque eget ligula. Fusce cursus justo at sem auctor dapibus. Proin eget quam sed orci eleifend molestie. Suspendisse malesuada velit eget lectus porttitor iaculis.

5.2 Raze the ground a confronto con NUOVO/I METODO/I DA TROVATE

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur non nunc in purus aliquam eleifend. Sed sed justo in nisl fringilla vehicula. Aenean sodales pellentesque porttitor. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Morbi lacus lacus, laoreet quis pretium non, dignissim sit amet purus. Vivamus eu mauris felis, eget vulputate metus. Vestibulum sollicitudin nisi vitae quam venenatis id ornare magna condimentum. Nullam faucibus commodo dui quis tempor. Donec sed placerat odio. Sed varius, mi et volutpat egestas, leo arcu laoreet ante, in commodo ipsum neque eget ligula. Fusce cursus justo at sem auctor dapibus. Proin eget quam sed orci eleifend molestie. Suspendisse malesuada velit eget lectus porttitor iaculis.

Conclusioni

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur non nunc in purus aliquam eleifend. Sed sed justo in nisl fringilla vehicula. Aenean sodales pellentesque porttitor. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Morbi lacus lacus, laoreet quis pretium non, dignissim sit amet purus. Vivamus eu mauris felis, eget vulputate metus. Vestibulum sollicitudin nisi vitae quam venenatis id ornare magna condimentum. Nullam faucibus commodo dui quis tempor. Donec sed placerat odio. Sed varius, mi et volutpat egestas, leo arcu laoreet ante, in commodo ipsum neque eget ligula. Fusce cursus justo at sem auctor dapibus. Proin eget quam sed orci eleifend molestie. Suspendisse malesuada velit eget lectus porttitor iaculis.

Sed eget ullamcorper ligula. Curabitur at massa at ante porta imperdiet. Proin neque est, bibendum facilisis faucibus sed, porttitor quis nisl. Praesent imperdiet gravida interdum. Fusce non odio neque, sit amet lobortis felis. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus quis erat eu elit aliquam eleifend ac ut ante. Vestibulum posuere placerat arcu, nec varius urna aliquam vitae. Phasellus dapibus eros nec eros laoreet sed bibendum lacus rhoncus. Nunc luctus sem sit amet leo hendrerit non imperdiet nisi pellentesque. Pellentesque iaculis odio sit amet sem ornare ultrices. Phasellus ante est, viverra non posuere eget, facilisis id velit. Vivamus accumsan eros vel magna cursus mollis at in odio. Ut a semper mauris. In turpis metus, lacinia a malesuada sed, faucibus eget lectus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Nunc malesuada aliquam urna in suscipit. Vivamus vel lacinia enim. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aenean purus ante, dignissim nec pulvinar quis, dignissim quis neque.

Suspendisse posuere adipiscing leo id pellentesque. Ut eros massa, viverra sit amet consequat id, molestie et nisl. Nunc ac venenatis quam. Quisque scelerisque, risus in pretium pharetra, leo orci egestas eros, eu pellentesque lacus massa vel tellus. Fusce adipiscing faucibus libero in tempus. Maecenas sed neque sed sapien tincidunt ornare ut et sem. Praesent vitae dui mauris,

vitae tristique lorem. Etiam dui odio, malesuada ultrices suscipit ut, vulputate rhoncus metus. Nulla faucibus fringilla fringilla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Pellentesque vel elit lectus, dictum placerat est. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Sed suscipit sagittis auctor. Phasellus adipiscing placerat varius. Sed sed sem lacus.

Ringraziamenti

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur non nunc in purus aliquam eleifend. Sed sed justo in nisl fringilla vehicula. Aenean sodales pellentesque porttitor. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Morbi lacus lacus, laoreet quis pretium non, dignissim sit amet purus. Vivamus eu mauris felis, eget vulputate metus. Vestibulum sollicitudin nisi vitae quam venenatis id ornare magna condimentum. Nullam faucibus commodo dui quis tempor. Donec sed placerat odio. Sed varius, mi et volutpat egestas, leo arcu laoreet ante, in commodo ipsum neque eget ligula. Fusce cursus justo at sem auctor dapibus. Proin eget quam sed orci eleifend molestie. Suspendisse malesuada velit eget lectus porttitor iaculis.

Sed eget ullamcorper ligula. Curabitur at massa at ante porta imperdiet. Proin neque est, bibendum facilisis faucibus sed, porttitor quis nisl. Praesent imperdiet gravida interdum. Fusce non odio neque, sit amet lobortis felis. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus quis erat eu elit aliquam eleifend ac ut ante. Vestibulum posuere placerat arcu, nec varius urna aliquam vitae. Phasellus dapibus eros nec eros laoreet sed bibendum lacus rhoncus. Nunc luctus sem sit amet leo hendrerit non imperdiet nisi pellentesque. Pellentesque iaculis odio sit amet sem ornare ultrices. Phasellus ante est, viverra non posuere eget, facilisis id velit. Vivamus accumsan eros vel magna cursus mollis at in odio. Ut a semper mauris. In turpis metus, lacinia a malesuada sed, faucibus eget lectus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Nunc malesuada aliquam urna in suscipit. Vivamus vel lacinia enim. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aenean purus ante, dignissim nec pulvinar quis, dignissim quis neque.

Suspendisse posuere adipiscing leo id pellentesque. Ut eros massa, viverra sit amet consequat id, molestie et nisl. Nunc ac venenatis quam. Quisque scelerisque, risus in pretium pharetra, leo orci egestas eros, eu pellentesque lacus massa vel tellus. Fusce adipiscing faucibus libero in tempus. Maecenas sed neque sed sapien tincidunt ornare ut et sem. Praesent vitae dui mauris,

vitae tristique lorem. Etiam dui odio, malesuada ultrices suscipit ut, vulputate rhoncus metus. Nulla faucibus fringilla fringilla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Pellentesque vel elit lectus, dictum placerat est. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Sed suscipit sagittis auctor. Phasellus adipiscing placerat varius. Sed sed sem lacus.

Elenco delle figure

Bibliografia

[1] G. Hinton, O. Vinyals e J. Dean. *Distilling the Knowledge in a Neural Network*. 2015. arXiv: 1503.02531 [stat.ML]. URL: https://arxiv.org/abs/1503.02531.

[2] W. Murray e K.-M. Ng. *An algorithm for nonlinear optimization problems with binary variables*. 2010. URL: https://link.springer.com/article/10.1007/s10589-008-9218-1.