

# Descripció de la pràctica

XNDL 2022-2023 Q2

El 40% de la nota final de l'assignatura s'obté d'aquest laboratori. Es realitza en parelles, l'enunciat s'entrega als estudiants el dia 3-4 de Maig, i s'haurà de lliurar com a màxim el 2 de Juny.

La pràctica consisteix de tres tasques independents. Cada parella ha de realitzar les tres tasques de manera col·laborativa.

Per garantir una dedicació comparable a les tasques, la nota final serà la mitja ponderada de les tres notes, on la pitjor comptarà el doble. Aquest document descriu cadascuna de les tres tasques, així com els criteris d'avaluació (i.e., rúbrica).

## 1. Processament d'imatge

Els estudiants hauran de entrenar un seguit de xarxes neuronals profundes per a solucionar un problema realista d'imatge. Els alumnes rebran un codi funcional amb un model subòptim. Basat en aquest, hauran de realitzar el següent cicle de raonament experimental:

1. Diagnòstic de la xarxa: Identificar quin model es pren com a punt de partida de l'iteració, i perquè. Descriure el comportament i rendiment de la xarxa. **Raonar les observacions en funció de les eines de diagnòstic** (corbes de pèrdua, mètriques, matrius de confusió, "*model summary*"). Entre els possibles diagnòstics estan l'underfit, fit, overfit, inestabilitat de gradients i de procés d'aprenentatge, manca de convergència, convergència excessivament ràpida/lenta, comportament aleatori, explosió de loss, etc.
2. Creació d'hipòtesis: En funció del diagnòstic, llistar les diferents metodologies disponibles per intentar millorar-ne el rendiment. Per exemple, si te overfit podem fer A, B, C o D. Dintre del llistat d'eines disponibles, **triar i justificar la tria d'una (i només una) alternativa tenint en compte les eines de diagnòstic**.
3. Disseny i execució d'experiments: **Descriure els experiments a realitzar per avaluar si la metodologia triada al punt (2) millora el model seguint el diagnòstic del punt (1)**. Per exemple, si triem incrementar la capacitat del model, podem provar a afegir 1 o 2 capes, amb un nombre X o Y de neurones. Si decidim afegir data augmentation, podem provar varies combinatòries de transformacions per trobar la més adient.

Aquest cicle es podrà realitzar un màxim de 7 vegades. *No és podrà canviar cap paràmetre o hyperparàmetre del model original fora d'aquest cicle*. Als models sortint d'aquest cicle (al final del punt (3)) se'ls ha d'identificar de manera única (e.g., *model\_zero* es el model punt d'origen donat als alumnes). Cada iteració (punt (1)) pot començar sobre qualsevol model prèviament generat (e.g., per si voleu canviar de estratègia, i tornar enrere a un model anterior). El document adjunt "Final lab reporting" conté un formulari guia per documentar aquest procés.

El problema a resoldre es el Museum Medium Art dataset (MAMe [1]) en la seva versió de baixa resolució. Usarem el train per entrenar, i el validation per a orientar cicle de raonament. El test s'haurà de correr al final de l'experimentació sobre el model que sigui

considerat el millor per part del estudiant. Com a baseline, es pot assumir que un bon model arriba al 70% de precisió, i un molt bon model al 80%.

### Rúbrica

<i>Criteri</i>	<i>Satisfactori (10-9)</i>	<i>Millorable (8-5)</i>	<i>Malament (4-0)</i>
<i>Diagnòstic (x7)</i> 2.5 punts	El diagnòstic del model <b>considera totes</b> les evidències disponibles, és <b>coherent</b> amb aquestes i esta adequadament <b>raonat</b> .	El diagnòstic té en compte algunes evidències pero <b>no totes</b> , o és <b>parcialment contradictori</b> amb aquestes o <b>no esta completament raonat</b> .	El diagnòstic <b>no té en compte les principals evidències</b> , és <b>incoherent</b> , o <b>no te raonament</b> .
<i>Hipòtesis (x7)</i> 2.5 punts	Les hipòtesis generades son <b>coherents amb el diagnòstic</b> , consideren <b>la gran majoria de les opcions</b> disponibles, i <b>la tria és perfectament consistent</b> amb el estat del model.	Les hipòtesis son <b>parcialment coherents</b> amb el diagnostic, o <b>consideren part de les opcions</b> disponibles, o la <b>tria esta relacionada</b> amb l'estat del model.	Les hipòtesis son <b>incoherents</b> amb el diagnostic, o no consideren una <b>porció rellevant de les opcions</b> disponibles, o la <b>tria és parcialment arbitraria</b> .
<i>Experiments (x7)</i> 2.5 punts	Els experiments son <b>complets pero no excessius</b> . <b>Permeten confirmar o descartar la millora del model</b> , i per tant la hipòtesis anterior.	Els <b>experiments son excessius o incomplets</b> . Podrien validar la hipòtesis, o no fer-ho.	Els experiments son <b>incoherents amb la hipòtesis</b> .
<i>Cicles i tria de model</i> 1.25 punts	S'han realitzat els 7 experiments de manera coherent. Triant com a model d'inici un que permet realitzar un cicle experimental de rellevància.	S'han realitzat més o menys de 7, o en alguns casos el model triat d'inici de cicle no es adient.	S'han realitzat menys de 4 cicles d'experiments, i en la majoria dels casos el model triat no es adient.
<i>Ús de les dades</i> 1.25 punts	S'han usat les tres particions de dades correctament.	Alguna de les particions no ha estat usada per al seu propòsit.	Les particions han estat usades de manera que els resultats i/o el model no son fiables.

A continuació, un exemple de cicle realitzat satisfactòriament:

- (1) Partim del últim model entrenat, model\_A2, ja que era el millor fins al moment. Observant les corbes de train i val loss i les mètriques adjuntes veiem que hi ha overfitting. El model summary mostra que el nostre model te 50M de paràmetres, fet que recolza aquesta observació. El train pero no pasa del 80%, l'aprenentatge és sorollós i triga a convergir.
- (2) Per reduir l'overfitting podem reduir la complexitat, afegir regularització, incrementar el batch size. Donat que el train no arriba al 80%, decidim no complicar més el problema (no regularització o data augmentation) i reduir la complexitat.
- (3) Tenint en compte que la majoria de la complexitat es troba en la connexió de la 3a i 4a capa, realitzarem els següents experiments:
  - (a) Reduir el nombre de neurones d'aquestes capes. Meitat de neurones en les primer 3 i 6 capes. (2 experiments)
  - (b) Afegir mètodes de reducció de dimensionalitat (pooling, stride) per limitar la complexitat d'aquestes capes sense canviar el nombre de neurones. Average Pooling (2x2) i stride (2) a una de cada dues capes convolucionals (2 experiments)
  - (c) Eliminar una capa de la xarxa. La última o la primera (2 experiments)

Total experiments: 6

Cladrà entregar el document “Final lab reporting” amb els detalls de l’experimentació.

[1] <https://hpaibsc.es/MAMe-dataset/>

## 2. Processament d’embeddings

Els embeddings son representacions intermitges apreses per les xarxes neuronals. Per una entrada donada, aquests vectors codifiquen la informació disponible per la xarxa. En els “word embeddings” el model es entrenat per aprendre la relació entre cada paraula i les altres paraules en el seu context (e.g., amb una finestra de 5 paraules). D’aquesta manera, podem explorar les relacions entre paraules i textos.

Per aquest laboratori es proporcionen dos codis and diferents funcionalitats. Ambós utilitzen els word embeddings de glove [2], i us ajudaran a realitzar dues tasques (A i B). El document adjunt “Final lab reporting” conté un formulari per documentar aquest procés:

- word\_regularities.py
  - Computar distàncies entre paraules
  - Cercar les paraules més similars a un terme
  - Calcular les paraules més properes a una “analogia” (X-Y+Z=?)

Heu de (A1) proposar dues regularitats originals, com les llistades a la Taula 1 (e.g., pais-capital, pais-moneda, masculí-femení, etc...). Per cadascuna llisteu cinc parelles de paraules que poden complir la regularitat. (A2) Dissenyeu els experiments i (A3) analitzeu els resultats.

Table 1: *Examples of five types of semantic and nine types of syntactic questions in the Semantic-Syntactic Word Relationship test set.*

Type of relationship	Word Pair 1		Word Pair 2	
Common capital city	Athens	Greece	Oslo	Norway
All capital cities	Astana	Kazakhstan	Harare	Zimbabwe
Currency	Angola	kwanza	Iran	rial
City-in-state	Chicago	Illinois	Stockton	California
Man-Woman	brother	sister	grandson	granddaughter
Adjective to adverb	apparent	apparently	rapid	rapidly
Opposite	possibly	impossibly	ethical	unethical
Comparative	great	greater	tough	tougher
Superlative	easy	easiest	lucky	luckiest
Present Participle	think	thinking	read	reading
Nationality adjective	Switzerland	Swiss	Cambodia	Cambodian
Past tense	walking	walked	swimming	swam
Plural nouns	mouse	mice	dollar	dollars
Plural verbs	work	works	speak	speaks

- doc\_analysis.py
  - Calcula l'embedding d'un text, calculant el vector mitja de totes les seves paraules (el llibre "The fellowship of the ring" partit en 3 parts).
  - Comparar l'embedding de cada part amb el de certs termes.

Per aquesta part (B) primer heu de triar un dataset. Poden ser llibres, guions, notícies, xarxes socials, wikipedia o qualsevol altre font de text (e.g., el llibre "the fellowship of the ring"). Si teniu dubtes sobre la font triada, consulteu al professor. Donat el text disponible, heu de (B1) generar una hipòtesis (e.g., "el llibre és un viatge de la felicitat a la violència"). (B2) Plantejar com la podeu testejar com a un problema de classificació o regressió i descriure la partició del dataset (e.g., puc predir a quin terç del llibre pertany un paragraf usant "paragraph embeddings", creant un dataset com el de la Figura 2.). (B3) dissenyar i entrenar un MLP per resoldre el problema.

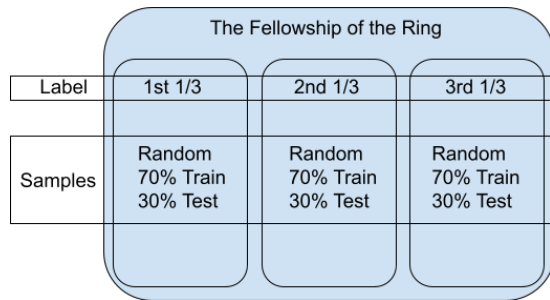


Figure 2

Criteri	Satisfactori (10-9)	Millorable (8-5)	Malament (4-0)
(A1) Hipòtesis 1.5 punts	Les dues regularitats son originals i raonables. Les dues llistes son completes i adients.	Una de les dues regularitats es original i raonable. I la seva llista es completa i adient.	Cap de les dues regularitats es original o raonable. Cap llista completa i adient.
(A2) Experiments 1.5 punts	Els experiments estan ben descrits i son suficients per validar o refutar hipòtesis.	Els experiments son parcialment incomplets, o insuficients validar o refutar la hipòtesis.	Els experiments estan pobrament presentats, o son incongruents amb l'hipòtesis.
(A3) Presentació de resultats 1.5 punts	L'anàlisi dels resultats és complert i d'acord amb l'evidència. Totes les afirmacions son validables amb la informació presentada.	L'Anàlisis dels resultats és incomplert o parcialment incongruent. Manca informació per sostindre les afirmacions realitzades.	Majoritàriament incorrecte o incongruent anàlisis de resultats. Gran manca d'informació sobre els resultats,
(B1) Hipòtesis 1.5 punts	L'Hipòtesis es original i raonable per el text escollit.	L'Hipòtesis no es del tot original o raonable per el text escollit.	L'Hipòtesis no es original o raonable per el text escollit.

(B2) <i>Experiments</i> 1.5 punts	Els experiments estan ben descrits i son suficients per validar o refutar hipòtesis.	Els experiments son parcialment incomplets, o insuficients validar o refutar l'hipòtesis.	Els experiments estan pobrament presentats, o son incongruents amb l'hipòtesis.
(B3) <i>MLP</i> 2.5 punts	Disseny i tria de hiperparàmetres adient i ben raonada per el problema.	Disseny i tria de hiperparàmetres deficient o incompletament raonada per el problema.	Disseny i tria de hiperparàmetres arbitraria o injustificada.

[2] <https://nlp.stanford.edu/projects/glove/>