

Beskrivelse af eksamensopgave

Den praktiske del af opgaven

* = Implementering findes på itslearning

** = Hjælp til implementering findes på itslearning

Assignment del 1:

- Arbejde med KNN til klassifikation af CIFAR-10 *
- Arbejde med SVM til klassifikation af CIFAR-10 *
- Arbejde med Softmax til klassifikation af CIFAR-10 *
- Arbejde med et 2-lags NN til klassifikation af CIFAR-10 *

Assignment del 2:

- I skal implementere et Convolutional Neural Network til klassifikation af CIFAR-10. I dette skal bl.a. indgå følgende:
 - Del data op i træning, validation og test
 - Load og præ-processér data
 - Afprøv forskellige netværksarkitekturer og hyperparametre således at den bedste performance opnås på validation-sættet.
 - Mulighed for at implementere data augmentation
 - Mulighed for at anvende transfer-learning
 - Test til sidst den trænede model med de bedste hyperparametre på test-sættet.
 - I kan bruge Keras med Tensorflow som backend (eller anden framework I vælger).

Assignment del 3:

I denne del af opgaven skal I vælge én af disse tre opgaver:

1. I skal implementere et Convolutional Neural Network til segmentering af et datasæt i selv vælger (f.eks. fra Kaggle, eller et datasæt I selv laver). I dette skal bl.a. indgå følgende:
 - Download eller lav selv et datasæt
 - Upload datasættet til Google Drive, så I kan loade det ind fra Google Colab
 - Del data op i træning, validation og test
 - Load og præ-processér data
 - Afprøv forskellige netværksarkitekturer** og hyperparametre således at den bedste performance opnås på validation-sættet.
 - I kan afprøve en eller flere forskellige loss funktioner.
 - Test til sidst den trænede model med de bedste hyperparametre på test-sættet.
 - Konvertér segmenterede test-billeder til RGB-format **
 - I kan bruge Keras med Tensorflow som backend (eller anden framework I vælger).
2. You should implement a deep learning based time-series classification algorithm. You can use one of the following health related datasets/time-series from the list below:
 - MIT-BIH Arrhythmia Database (<https://physionet.org/content/mitdb/>)
 - MIT-BIH Atrial Fibrillation Database (<https://physionet.org/content/afdb/>)
 - AF Classification from a Short Single Lead ECG Recording - The PhysioNet Computing in Cardiology Challenge 2017 (<https://physionet.org/content/challenge-2017/>)
 - Classification of 12-lead ECGs: the PhysioNet - Computing in Cardiology Challenge 2020 (<https://physionet.org/content/challenge-2020/>)

You can follow the below steps to complete the task:

- Use the above links to download the dataset. You can use the PhysioNet WFDB python package to download the entire dataset easily. Please see the demo.ipynb file on this GitHub page (<https://github.com/MIT-LCP/wfdb-python>).

- Upload the dataset to Google Drive, so that you can load it from Google Colab using the WFDB library.
 - Split the data into train, validation, and test sets.
 - Load and preprocess the data e.g. filter noise and remove baseline wander of ECG signals
 - Train different networks architectures** and hyperparameters to achieve the best result on validation set. This can be a LSTM, CNN, or a combination of both models.
 - The performance of the best trained model with its optimum hyperparameters should be evaluated on the test set.
 - You should report the results of the model using different metrics and techniques such as confusion matrix, Receiver Operating Characteristics (ROC) curve, sensitivity, precision, etc.
3. I skal implementere en tekstklassificeringsalgoritme ved brug af en convolutional eller recurrent arkitektur med et datasæt I selv vælger. Det kunne fx være sentiment klassificering.
- Download et datasæt (fx IMDB datasæt)
 - Upload datasættet til Google Drive, så I kan loade det ind fra Google Colab
 - Del data op i træning, validation og test
 - Foretag eksplorativ dataanalyse
 - Præ-processér data. Eksperimentér fx med stop-words, stemming, lemmatization, etc.
 - Afprøv forskellige netværksarkitekturer** og hyperparametre således at den bedste performance opnås på validation-sættet. Det kunne fx være RNN, GRU, LSTM eller CNN arkitekturer. Prøv evt. også med og uden præ-trænede word embeddings.
 - Test til sidst den trænede model med de bedste hyperparametre på test-sættet.
 - Lav en confusion matrix og analyser resultatet.
 -

Arbejdet dokumenteres i en samlet rapport

Rapporten skal som minimum indeholde følgende punkter:

Assignment del 1:

- Kort beskrivelse af problemet (CIFAR-10 billede klassifikation)
 - Beskrivelse af datasættet (klasser, billedstørrelse, m.m.)
 - Beskrivelse af udfordringer ved problemet (f.eks. semantic gap)
 - Train/val/test split
- Kort teoretisk beskrivelse af KNN, SVM, Softmax & NN
 - Relevante formler & figurer
- Hvordan har I trænet dem?
- Hvilke hyperparametre har I testet? Hvorfor?
- Resultater (på både træning-, validation- og testsæt!)

Assignment del 2:

- Dybdgående teoretisk beskrivelse af Convolutional Neural Networks til klassifikation
- Beskrivelse af og argumentation for netværksarkitektur
- Hvilke eksperimenter har I lavet? (Forskellige netværksarkitekturer, antal filtre, tuning af hyperparametre m.v.)
- Resultater (på både træning-, validation- og testsæt!)
- Diskussion
 - Hvad virkede, hvad virkede ikke og hvorfor, hvad kunne man ellers prøve?
 - Hvilken effekt havde forskellige hyperparametre?
- Konklusion

Assignment del 3:

Hvis I valgte segmenteringsopgaven:

- Hvordan adskiller segmentering sig fra klassifikation? Hvordan er netværkene anderledes?
- Beskrivelse af og argumentation for netværksarkitektur
- Hvilke eksperimenter har I lavet? (Forskellige netværksarkitekturer, antal filtre, tuning af hyperparametre m.v.)
- Resultater (på både træning-, validation- og testsæt!)
- Diskussion
 - Hvad virkede, hvad virkede ikke og hvorfor, hvad kunne man ellers prøve?
 - Hvilken effekt havde forskellige hyperparametre?
- Konklusion

If you choose the time-series analysis task:

- Abstract
 - enough information should be provided get a clear idea about the subject matter.
- Introduction
 - required background information to understand the work should be covered
 - the motivation why this project is important and interesting should be given.
- Related Work
 - similar works should be discussed
 - references should be cited properly
 - the proposed method should be compared wither state-of-the-arts.
- Proposed Method
 - describe the main algorithm so that it can be implemented by others.
- Experiments

- experimental setup and methodology should be described well enough so that it can be repeated by others
- describe the used dataset appropriately.
- Results and Discussion
 - the results should be described clearly
 - the data should also be analyzed well
 - the figures should be clear and have no missing labels
 - make sure that figure captions have sufficient information to understand the figure
 - each figure should be also referenced in the text
 - the discussion should be critical/honest, and potential weaknesses/shortcomings should be discussed as well.

Hvis I valgte opgaven om natural language processing:

- Teori
 - Beskriv teorien for word embeddings
 - Beskriv teorien for recurrent og convolutional arkitekturer og fordele/ulemper ved de forskellige
- Metode
 - Beskriv det valgte datasæt og klassificeringsproblem
 - Beskriv og argumenter for din valgte præ-processering
 - Argumenter for dit valg af word embeddings (bruger du præ-trænede word embeddings og/eller trænes de sammen med modellen) og beskriv teorien bag
 - Argumenter for dit valg af netværkstype (convolutional eller recurrent) og netværksarkitektur
 - Argumenter for de valgte hyperparametre
- Resultater
 - Beskriv resultatet
- Diskussion
 - Analyser resultatet - hvilke fejl laver netværket?
- Konklusion
 - Opsummer det vigtigste resultat

Rapporten skal som minimum inkludere grafer som fx. loss-kurver og accuracy-kurver. Disse kan også vedlægges i appendix.

Kode/kodeeksempler inkluderes ikke i rapporten, men kan leveres i appendix, hvis behovet opstår.

Omfang ~ 25 sider

Aflevering: **19. december kl. 23:59**

OBS: You are NOT allowed to write your personal identification number (CPR) on your hand-in. / Bemærk, at du IKKE må skrive cpr-nummer på din aflevering.