

Kunskapskontroll 2

DS24 Deep Learning

Kunskapskontrollen består av två delar: teoretiska frågor samt ett tillämpat projekt kopplat till chattbottar.

Del 1: Teoretiska frågor

1. Hur är AI, Maskininlärning och Deep Learning relaterat?
2. Hur är Tensorflow och Keras relaterat?
3. Vad är en parameter? Vad är en hyperparameter?
4. När man skall göra modellval och modellutvärdering kan man använda tränings-, validerings- och testdataset. Förklara hur de olika delarna kan användas.
5. Förklara vad nedanstående kod gör:

```
n_cols = x_train.shape[1]

nn_model = Sequential()
nn_model.add(Dense(100, activation='relu', input_shape=(n_cols, )))
nn_model.add(Dropout(rate=0.2))
nn_model.add(Dense(50, activation='relu'))
nn_model.add(Dense(1, activation='sigmoid'))

nn_model.compile(
    optimizer='adam',
    loss='binary_crossentropy',
    metrics=['accuracy' ])

early_stopping_monitor = EarlyStopping(patience=5)
nn_model.fit(
    x_train,
    y_train,
    validation_split=0.2,
    epochs=100,
    callbacks=[early_stopping_monitor])
```

6. Vad är syftet med att regularisera en modell?
7. "Dropout" är en regulariseringsteknik, vad är det för något?
8. "Early stopping" är en regulariseringsteknik, vad är det för något?
9. Din kollega frågar dig vilken typ av neuralt nätverk som är populärt för bildanalys, vad svarar du?
10. Förklara översiktligt hur ett "Convolutional Neural Network" fungerar.
11. Vad gör nedanstående kod?

```
model.save("model_file.keras")  
my_model = load_model("model_file.keras")
```

12. Deep Learning modeller kan ta lång tid att träna, då kan GPU via t.ex. Google Colab skynda på träningen avsevärt. Skriv mycket kortfattat vad CPU och GPU är.

Del 2: Implementera en chattbot med RAG

Den här delen går ut på att skapa en chattbot som använder RAG-teknik för att begränsa svaren den ger till en viss kontext. Du väljer själv vilket användningsområde och vilka dokument du ska använda. Det kan vara en viss bok, en uppsättning dokument kopplade till en organisation eller ett företag, eller liknande. Det är också upp till dig om du vill använda en molnbaserad språkmodell som grund, eller om du vill använda en lokal språkmodell. Chattboten kan antingen köras som en Streamlit-applikation, eller textbaserat i terminalen.

Lämna in din kod i Omniway via en GitHub-länk (Gör din repo publik!). Arbetet lämnas in som kodfiler där koden ska vara kommenterad. Detta är den "skriftliga rapportdelen" i kunskapskontrollen. I slutet av din kod ska du redogöra för hur din modell hade kunnat användas i verkligheten och vilka potentiella utmaningar och möjligheter (t.ex. affärsmässiga, etiska och andra perspektiv du finner relevanta) som finns. Du kan skriva detta som kommentarer i koden.

Betygsättning

Godkänt

- Du har gjort en chattbot som använder sig av RAG-teknik för att begränsa svaren till en viss kontext.

Väl godkänt

- Du har gjort en chattbot som använder sig av RAG-teknik för att begränsa svaren till en viss kontext.
- Du har också gjort ett system för att evaluera chattboten.
- Välskriven och korrekt kod.

- En fördjupad och kritisk diskussion kring *“hur din modell hade kunnat användas i verkligheten och vilka potentiella utmaningar och möjligheter (affärsmässiga, etiska och andra perspektiv du finner relevanta) som finns”*.
- Tydliga och koncisa svar på de teoretiska frågorna.

Självutvärdering

Du ska också lämna in en självutvärdering där du svarar på följande frågor:

1. Vad har varit roligast i kunskapskontrollen?
2. Vilket betyg anser du att du ska ha och varför?
3. Vad har varit mest utmanande i arbetet och hur har du hanterat det?

Deadline

Sista dag för inlämning är fredag 30/5 kl. 17. Det går att lämna in fram till söndag 1/6 men det är generellt bra att hålla gängse kontorstider.