# Artificiell Intelligens - Kunskapskontroll 2 Del 1

Namn: Björn Axelsson

Kunskapskontrollen lämnas in individuellt via en GitHub mapp men ni uppmuntras till att samarbeta och använda alla hjälpmedel så länge som ni förstår vad ni gjort och det ni lämnar in är ert egna arbeta, precis som i arbetslivet.

Kunskapskontrollen har en VG del som görs av de som satsar på betyget VG.

## G-delen

G-delen består av att ni skall (1) besvara teoretiska frågor och (2) modellera MNIST datan. Teoretiska frågor

Besvara nedanstående teoretiska frågor koncist.

1. Lotta delar upp sin data i ”Träning”, ”Validering” och ”Test”, vad används respektive del för?

**Träningsdata** – data som används för att träna modeller

**Valideringsdata** – data som används för att utvärdera modeller där man väljer den bästa

**Testdata** – data som används för att testa den valda modellen på ny data för att få en uppskattning på modellens generaliseringsförmåga.

1. Förklara (gärna med ett exempel): Ordinal encoding, one-hot encoding, dummy variable encoding.

**Ordinal encoding** – kategorier med en naturlig inbördes ordning eller rangordning, exempelvis utbildningsnivå, betyg eller kundnöjdhet.

**One-hot encoding** – kategorier utan naturlig inbördes ordning eller rangordning, exempelvis kön eller färg

**Dummy variable encoding** – liknar one-hot encoding men kategorier exkluderas för att undvika miltikollinearitet, används vid linjär regression. Exempel på användning kan vara färg eller kategoriska grupper.

1. Göran påstår att datan antingen är ”ordinal” eller ”nominal”. Julia säger att detta måste tolkas. Hon ger ett exempel med att färger såsom {röd, grön, blå} generellt sett inte har någon inbördes ordning (nominal) men om du har en röd skjorta så är du vackrast på festen (ordinal) – vem har rätt?

Båda har rätt, något är antingen ordinal eller nominal men det behöver tolkas utifrån situationen och det dataset som används. Så utifrån det skulle jag säga att Julia har kommit lite längre i sin förståelse än Göran som verkar ha ett mer binärt förhållningssätt. Färger har inte någon inbördes ordning men det beror som Julia beskriver på situationen, om färgen röd gör dig vackrare så får den färgen ett annat värde än de andra färgerna och i det sammanhanget så blir färger i så fall ordinal.

1. Läs följande länk: https://stackoverflow.com/questions/56107259/how-to-save-a-trained-model-by-scikit-learn (speciellt svaret från användaren som heter ”sentence”) som beskriver ”joblib” och ”pickle”. Det är alltså ett sätt att spara modeller och innebär att man kan träna en modell och sedan återanvända den för att göra prediktioner utan att behöva träna om modellen. Detta kommer ni ha nytta av om ni satsar på VG delen. Svara på frågan: Vad används joblib och pickle till?

Istället för att köra om hela modellen vilket kan ta timmar så sparar du modellen i en fil ”model.pkl” som sedan kan användas när du kör en predict på ny data.

## Modellera MNIST

Använd maskininlärning för att modellera MNIST datan. Du skall utvärdera minst två olika modeller i ditt arbete och göra ett komplett ML-flöde, från början, där du laddar in data, till slut, där du utvärderar den bäst valda modellen på din test data. Hur du laddar ned MNIST datan kan du se här.

<https://github.com/Mrcolgate2024/NBI_python_delkurs2_kunskapskontroll>

<https://github.com/Mrcolgate2024/NBI_python_delkurs2_kunskapskontroll/blob/main/mnist_ensemble_kunskapskontroll2_del1_FINAL.ipynb>

## VG delen

De som satsar på VG behöver genomföra arbetet i G delen på ett bra sätt samt göra VG delen nedan där en Streamlit applikation ska skapas.

* Skapa funktionalitet så att man kan göra *ett* av följande tre alternativ:
* Ladda upp en bild till Streamlit appen som predikterar vilken siffra det är.
* Rita en siffra på en ”rit-del” i Streamlit där siffran man ritat predikteras.
* Kunna ta bilder via datorkameran som sedan predikterar vilken siffra det är.

Notera: Du kan spara din modell från G-delen (t.ex. via joblib eller pickle) och använda den i denna delen.

**Notera, generellt sett är det stora jobbet att man behöver transformera bilderna som tas innan de används för att skapa en prediktion. Googla och diskutera med klasskamrater för att hantera problem som uppstår.**

LOCAL REPO: <https://github.com/Mrcolgate2024/NBI_python_delkurs2_kunskapskontroll>

STREAMLIT CLOUD REPO: <https://github.com/Mrcolgate2024/DigitSense>

PUBLIC LINK: <https://digitsense.streamlit.app/>