# Frågor – Kategorisk Data

Läs länken: <https://machinelearningmastery.com/one-hot-encoding-for-categorical-data/>

I Scikit-learn måste all data konverteras till numeriska variabler för att kunna användas. Besvara nedanstående instuderingsfrågor efter att ha läst länken ovan:

1. **Vad kännetecknar en nominell variabel?**

Svar: Variabeln kan anta ett diskret antal värden där det inte finns någon inbördes ordning, t.ex. {katt, hund, kanin}.

1. **Vad kännetecknar en ordinal variabel?**

Svar: Variabeln kan anta ett diskret antal värden där det finns en inbördes ordning, t.ex. {bäst, medel, sämst}.

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

Svar: I korta ordalag:   
- Ordinal encoding skapar en ny kolumn för en kategorisk variabel och numrerar kategorierna enligt inbördes ordning (startar ofta numreringen från 0).   
- One-hot-encoding skapar en ny kolumn för varje unik kategori i en feature (kolumn).

- Dummy variable encoding skapar (k-1) nya kolumner om det finns k kategorier i en feature (kolumn).

1. **Om vi använder vanlig linjär regression, skall vi använda one-hot-encoding eller dummy-variable encoding?**

Svar: Dummy variable encoding, anledningen är matematiskt teknisk och kan ignoreras.