# 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: Variabel som består av en ändlig kategorisk uppsättning diskreta värden utan relation mellan värden.

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

Svar: Variabel som består av en ändlig uppsättning diskreta värden med en rangordnad ordning mellan värden.

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

Svar:

* **Ordinal encoding**: Tex färger: [’R’,’G’,’B’] klassas om till [1,2,3]  
  ’R’ 1  
  ’R’ --\ 1  
  ’G’ --/ 2   
  ’B’ 3
* **One-hot-encoding**:

r g b  
---------- ----------------

’R’ 1 0 0  
’R’ --\ 1 0 0  
’G’ --/ 0 1 0  
’B’ 0 0 1

* **Dummy-variable-encoding**:  
  ---------- ----------------

’R’ 0 0  
’R’ --\ 0 0  
’G’ --/ 1 0  
’B’ 0 1

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

Svar: Dummy-variable-encoding