

Inlämningsuppgift 1 - Decision tree
Maskininlärning och Deep learning - VT 2022
Individuell inlämning
Lämna in senast 25/4 08:00

I denna inlämning ska kvaliteten på viner klassificeras med hjälp av decision trees.

Inlämningen är uppdelad i en godkänt del och en väl godkänt del.

För godkänt krävs godkänt på godkänt delen, inklusive efter komplettering.

För väl godkänt krävs att man blir godkänd på både godkänt delen och väl godkänt delen vid första inlämningen.

En notebook ska lämnas in som innehåller kod samt svaret på frågor. Tänk på att förklara bilder och svar. Döp filen till *för_efternamn_decision_tree.ipynb*

För hela uppgiften skall *WhineQT.csv* användas.

Datasetet innehåller information om vin samt kvaliteten på vinet (högre siffra - bättre kvalitet).

Godkänt:

Följande punkter skall behandlas och besvaras

- Analysera datan
 - Hur många datapunkter och attribut finns?
 - Hur många klasser av kvalitet finns det?
 - Hur ser distributionen ut för de olika attributen?
 - Finns det något attribut som inte borde användas?
- Dela upp datan i träning och testset
- Träna ett decision tree
- Vad får modellen för testscore?
- Hur djupt är trädet?
- Hur många löv har trädet?
- Vilka attribut har störst påverkan?

Väl Godkänt:

Följande punkter skall behandlas och besvaras

- Jämför träningsscore och testscore.
 - Hur skiljer de sig och varför?
 - Finns det något att göra åt fenomenet?
- Rita ut trädet
 - Hur många grenar kommer det maximalt från en nod?
 - Varför är det max så många grenar från en nod?

Utmaning (krävs ej för godkänt eller väl godkänt):

Kan du skapa en modell som har >0.7 i testscore? Om inte, vad är den bästa score du kan få?