

Inlämningsuppgift 2 - Random forest
Maskininlärning och Deep learning - VT 2022
Individuell inlämning
Lämna in senast 2/5 08:00

I denna inlämning ska kvaliteten på viner klassificeras med hjälp av random forest.

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_random_forest.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

- Träna en random forest
- Vad får modellen för testscore?
- Hur många träd finns det i skogen?
- Vilka attribut har störst påverkan?
- Vilka klasser är det den modellen har störst problem med att klassificera rätt? Varför är det problem med de klasserna?

Väl Godkänt:

Följande punkter skall behandlas och besvaras

- Bygg en egen random forest genom att träna flera decision trees.
 - Tänk på hur du ska jobba med träningsdatan för att få olika träd och att testdata inte får läcka till något träd.
 - Hur många träd har du i skogen?
 - Vad får du för testscore? Skiljer den sig mot modellen du skapade i godkänt delen? I så fall, varför tror du den skiljer sig?
 - Vilka klasser har modellen problem med? Är det samma klasser som för modellen i godkänt delen?

Utmaning (krävs ej för godkänt eller väl godkänt):

Vad är den bästa score du kan få, både för random forest för sklearn och för din egenbyggda skog?