Kunskapskontroll Sandra Persson

1. Teoretiska Frågor

1. Hur kan vi definiera ”Maskininlärning”?

2. Vad är supervised learning?

3. Vad är skillnaden mellan Regressionsproblem och Klassificieringsproblem?

4. Kan du ge ett exempel på vad regressionsmodeller respektive klassificeringsmodeller kan

användas till?

5. Vad är Root Mean Squared Error (RMSE)?

6. Vad är en ”confusion matrix”?

7. Om man delar upp datan i träning, validering och test – hur används respektive del?

8. Vad är en parameter för något? Ge ett exempel.

9. Vad är en hyperparameter för något? Ge ett exempel.

10. Vad är GridSearchCV i Scikit-learn? Ge ett exempel på vad det kan användas till.

2. Rapport

En ”grov mall” för hur rapporten skall vara strukturerad kan ni se nedan.

Introduktion som innehåller underrubrikerna: - Bakgrund

- Syfte och Frågeställning

Databeskrivning / EDA (Exploratory Data Analysis)

Metod och Modeller (Teori)

Projekt Resultat och Analys

- Resultat är deskriptiva i sin natur, t.ex. att man presenterar RMSE för sina olika modeller. Ofta kan tabeller vara användbara vid redogörelse av resultaten.

1

5. Slutsats och förslag på potentiell vidareutveckling.  
**Rapporten skall vara ca 2-3 sidor**. Skriv koncist och fundera på vad du vill lyfta fram.

**Efter att du är klar skall du skriva en kort redogörelse i slutet av rapporten (detta ingår inte i de 2-3 sidorna):**

**Utmaningar du haft under arbetet samt hur du hanterat dem.**

**Vilket betyg du anser att du skall ha och varför.**

**Tips du hade ”gett till dig själv” i början av kursen nu när du slutfört den.**

Introduktion

Bakgrund

Dataset housing, från Handson ML, från början ett dataset från StatLib.

10 variabler, 20 640 observationer.

9 numeriska och 1 kategorisk.

Longitude: interval scale

Latitude: interval scale

housing\_median\_age: ratio, discrete

total\_rooms: ratio, discrete

total\_bedrooms: ratio, discrete

population: ratio, discrete

households: ratio, discrete

median\_income: ratio

median\_house\_value: median house value for households within a block, in US dollars.

ocean\_proximity: kategorisk variabel,

Syfte och frågeställning

Databeskrivning/EDA

Metod och modeller

Resultat och analys

Slutsats och förslag på potentiell vidareutveckling