

Frågor till Föreläsning 4

1. Vad är scikit-learn?
2. Vilka steg finns för att bygga en Machine Learning modell?
 - a. Beskriv varje steg
3. Hur ser den generella formuleringen ut för Multiple Linear regression?
4. Hur tolkar man koefficienten?
5. Vad står "w" för?
6. Hur får man ut koefficienterna i scikit-learn?
7. Vad kallas den första koefficienten?
8. Hur får man ut den i scikit-learn?
9. Vad ska man tänka på vid Multiple Linear regression?
10. Vilka antaganden gör man när man använder Multiple Linear regression?
11. Hur evaluerar man en modell?
12. Vad menas med Mean Squared Error?

Övning

- Crime datasetet. Gör en multiple linear regression model med alla features inkludera. Jämför error metrics med modellen med 10 features från lektionen. Kan ni visualisera detta i samma plot?
- Välj ut top 20 features som är bäst korrelerade med target. Undersök sedan om de features är korrelerade med varandra. Gör ett urval av features från detta där ni väljer bort features som är korrelerade med andra (är två features korrelerade vill ni inte ta bort båda). Träna en multiple linear regression modell och jämför error metrics med de andra modellerna

Länkar

- Multiple linear regression <https://statsandr.com/blog/multiple-linear-regression-made-simple/>
- Ordinary least square <https://towardsdatascience.com/understanding-the-ols-method-for-simple-linear-regression-e0a4e8f692cc>
- Error metrics <https://machinelearningmastery.com/regression-metrics-for-machine-learning/>
- Skillnad MAE och RMSE <https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-better-e60ac3bde13d>