

## Föreläsning 06

### Övningsuppgifter

Prediktera Iris flower data. Använd datan iris.csv. Ni har tidigare utforskat datan. Målet med datan är att bestämma vilken typ av blomma det är, vi har alltså ett klassificeringsproblem. Nu ska ni prediktera med hjälp av Logistic regression.

### Frågor

1. Vad menas med Logic Regression
2. Vilka klasser av klassificering finns det
3. Vilken använder vi
4. Hur används den
5. Vad menas med Probability-Based Classifiers
6. Hur ser den generella formeln ut, samt vad står "w" och "x" för
7. Vad menas med positiv viktning
8. Vad menas med negativ viktning
9. Var går gränsen mellan positiv och negativ klass
10. Vad används evalueringsmetoder till
11. Rita upp ett träd med de evalueringsmetoder som finns
12. Vad är holdout cross-validation
13. Vad innebär confusion matrix
14. Rita upp en confusion mtrix
15. Milka "metrics of the confusion matrix" finns det samt definiera dem
16. Ge exempel på vilken metric man ska använda i ett visst läge
17. Vad är Threshold Value

### Data Camp

Supervised learning with scikit-learn <https://app.datacamp.com/learn/courses/supervised-learning-with-scikit-learn>

### Länkar

- Logistic regression (avancerat) <https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc>
- Dummy variables (avancerat) [https://www.sagepub.com/sites/default/files/upm-binaries/21120\\_Chapter\\_7.pdf](https://www.sagepub.com/sites/default/files/upm-binaries/21120_Chapter_7.pdf)