

Föreläsning 10

Övning

Ensemble methods for classification

- Ni ska jämföra Bagging, Random Forest och Boosting av ett klassificeringsproblem
- Datan ni ska använda är `credit_card_default.csv`. Kolla föreläsning 8 eller 9 för preprocessing av datan
- Jämföra Bagging, Random Forest och Boosting med Logistic Regression
- Följ samma tillvägagångssätt som demonstrerad i lektionen i `EnsembleMethodsRegression.ipynb`, fast anpassad för ett klassificeringsproblem. Det är alltså andra error metrics ni vill ta fram och jämföra
- Vilken metod fungerade bäst?
- Undersöka för bäst presterande modell vad som är optimal threshold genom att använda `precision_recall_curve` (föreläsning 8)
- Kolla ta reda på vilka features som Random Forest tycker är viktigast för predikteringen. Använd `feature_importances_`. Sortera och visualisera i en barplot

Frågor Ensemble methods

1. Kärnidén med denna metod är att anpassa en sekvens av svaga modeller och sedan i varje iteration vikta om datapunkterna från träningsdatasetet?
 - a. Random Forest
 - b. AdaBoost
 - c. Bagging
 - d. Decision Tree
2. Huvudkonceptet med ensemble methods är att?
 - a. Skapa många modeller och jämföra de
 - b. Kombinera prediktionerna till svaga modeller
 - c. Producera många modeller och välja den bästa
 - d. Kombinera prediktioner från många individuella prediktioner för att öka prestanda
3. I princip kan vilken modell som helst användas i Bagging?
 - a. Sant
 - b. Falskt
4. När du använder klassen `RandomForestClassifier` från `sklearn.ensemble` motsvarar parametern `n_estimators`?
 - a. Antal estimatorer i modellen
 - b. Antal träd i modellen
 - c. Antal parametrar att estimera
 - d. Antal estimatorer i varje iteration

Länkar

- Ensemble methods video https://www.youtube.com/watch?v=WtWxOhhZWX0&ab_channel=Simplilearn
- Ensemble methods scikit-learn <https://scikit-learn.org/stable/modules/ensemble.html>
- Ensemble methods <https://machinelearningmastery.com/tour-of-ensemble-learning-algorithms/>
- Natural Language Processing - Bag of Words <https://www.mygreatlearning.com/blog/bag-of-words/>

Data Camp

- Ensemble methods in Python <https://app.datacamp.com/learn/courses/ensemble-methods-in-python>
- Introduction to Natural Language Processing I Python <https://app.datacamp.com/learn/courses/introduction-to-natural-language-processing-in-python>