

Föreläsning 11

Övning

- Kolla på Credit card default datan. Jämför de fyra modellerna med cross validation och välj den bästa så som visad i lektionen.
- Läs om Chi-squared test <https://towardsdatascience.com/chi-square-test-for-feature-selection-in-machine-learning-206b1f0b8223>
- Läs om ANOVA test <https://towardsdatascience.com/anova-for-feature-selection-in-machine-learning-d9305e228476n>

Cross validation och hyperparameter tuning

1. Hur många uppskattningar av error metrics kan du få när du använder K-fold cross validation?
 - a. 1
 - b. K-1
 - c. K
 - d. K+1
2. Förutom att få en mer exakt uppskattning av error metrics, vad är K-fold cross validation användbar för?
 - a. Hyperparameter tuning
 - b. Få fler samples
 - c. Bättre visualisera modellen
 - d. Förbättra error metrics
3. Vad är hyperparametrar?
 - a. Parametrar som inte är direkt lärt inom modellen
 - b. Parametrar av meta-modellen
 - c. Parametrar som inte kan justeras
 - d. Ingen av ovanstående
4. Vilken scikit-learn klass används för att se efter bästa parametrar från en given grid genom att använda cross validation?
 - a. BaseSearchCV
 - b. SearchCV
 - c. CrossValidationSearch
 - d. GridSearchCV

Feature Selection

5. Vad är problemet med dummy features med låg varietet (varians)?
 - a. De ändras inte
 - b. De har sannolikt inte stor påverkan på prediktionen
 - c. De går modellen långsammare att träna
 - d. De skapar noisy features
6. Vilken statistisk test kan användas inom klassificering för att testa förhållandet mellan target och en kategorisk feature?
 - a. ANOVA-test
 - b. Fisher test
 - c. Chi-squared test
 - d. Pearson test

Länkar

- K-fold Cross Validation video https://www.youtube.com/watch?v=gJoOuNL-5Qw&ab_channel=codebasics
- K-fold Cross validation <https://towardsdatascience.com/k-fold-cross-validation-explained-in-plain-english-659e33c0bc0>
- Hyperparameter tuning <https://machinelearningmastery.com/hyperparameter-optimization-with-random-search-and-grid-search/>
- Hyperparameter tuning https://scikit-learn.org/stable/modules/grid_search.html
- Feature Selection ANOVA <https://towardsdatascience.com/anova-for-feature-selection-in-machine-learning-d9305e228476>
- Feature Selection Chi-Squared <https://towardsdatascience.com/chi-square-test-for-feature-selection-in-machine-learning-206b1f0b8223>

Data Camp

- Hyperparameter tuning <https://app.datacamp.com/learn/courses/hyperparameter-tuning-in-python>