



Data science i praktiken: algoritmer för maskininläring

Data science in practice: algorithms for machine learning

5 högskolepoäng

5 credits

Ladokkod: C2DP1C

Version: 2.2

Fastställt av: Utskottet för utbildningar inom bibliotek, information och IT 2022-03-24

Gäller från: HT 2022

Nivå: Avancerad nivå

Huvudområde (successiv fördjupning): Informatik (A1N)

Utbildningsområde: Naturvetenskap

Ämnesgrupp: Informatik/Data- och systemvetenskap

Förkunskapskrav: 120 högskolepoäng innehållande eller kompletterat med 15 högskolepoäng programmering.

Betygsskala: Underkänd eller Godkänd

Innehåll

Kursen syftar till att skapa en god förståelse för de olika algoritmer och tekniker som kan användas för maskininläring. Kursen introducerar den övergripande kontexten för maskininläring, med utgångspunkt i etablerade processer för informationsutvinning (CRISP-DM), samt teoribildning och terminologi inom maskininläring. Kursens huvudsakliga innehåll är grundläggande maskininläringstekniker som beslutsträd, regelinläring, neurala och djupa nätverk, support vector-maskiner och ensembler, samt algoritmer som kNN, k-Means och Apriori.

Kursens teoretiska innehåll behandlas i föreläsningar, med kompletterande seminarier som bidrar med forskningsanknytning, diskussion och reflektion. En betydande del av kursen utgörs av ett implementeringsprojekt där studenterna ges möjlighet till egen fördjupning inom något delområde av maskininläring. Exempelvis kan projektet behandla en specifik tillämpning, en vanligt förekommande uppgiftstyp (som klassificering, tidsserieanalys eller market basket analysis), eller någon teknik eller algoritm. I projektet läggs stor vikt vid systematiskt arbetssätt i experimentplanering och utvärdering.

Mål

Efter avslutad kurs ska studenten kunna, med avseende på:

Kunskap och förståelse

- 1.1 Visa förståelse för teori och terminologi inom maskininläring.
- 1.2 Visa god förståelse för centrala maskininläringstekniker och algoritmer.
- 1.3 Visa god förståelse för algoritmer och tekniker inom valda fördjupningsområden inom maskininläring.

Färdighet och förmåga

- 2.1 Använda och anpassa verktyg och tekniker från öppna standardramverk för att konstruera prototyper för maskininläring.
- 2.2 Planera och genomföra experiment för maskininläring med systematisk utvärdering.
- 2.3 Analysera algoritmers och modellers prestanda utifrån resultat av experiment.

Värderingsförmåga och förhållningssätt

- 3.1 Välja och argumentera för val av maskininlärningsalgoritm eller -teknik för ett givet problem i en verksamhetskontext.
- 3.2. Välja och värdera parameterkonfiguration av en maskininlärningsalgoritm i förhållande till en given frågeställning.
- 3.3 Diskutera och argumentera kring algoritmer och tekniker inom valda fördjupningsområden inom maskininläring.

Undervisningsformer

Undervisningen sker i form av:

- föreläsningar
- handledning

- seminarier

Undervisningen bedrivs på svenska, men undervisning på engelska kan förekomma.

Examinationsformer

Kursen examineras genom följande examinationsmoment:

Seminarier: Forskningsfördjupning

Lärandemål: 1.1 – 1.3, 2.2, 3.1 och 3.3

Högskolepoäng: 2,0

Betygsskala: Underkänd eller Godkänd

Projektuppgift: Individuell fördjupning med implementering och experiment inom maskininläring

Lärandemål: 1.1 – 1.3, 2.1 - 2.3 och 3.1 – 3.2

Högskolepoäng: 3,0

Betygsskala: Underkänd eller Godkänd

För betyget Godkänd på hel kurs krävs betyget Godkänd på samtliga examinationsmoment.

Seminarier kan ersättas med annan examinationsform om studenten underkänts på seminarium eller inte varit närvarande vid seminarium under kursens gång.

Om studenten har ett beslut/rekommendation om särskilt pedagogiskt stöd från Högskolan i Borås på grund av funktionsnedsättning, har examinator rätt att anpassa examinationen. Examinator har att utifrån kursplanens mål avgöra om examinationen kan anpassas i enlighet med beslutet/rekommendationen.

Studentens rättigheter och skyldigheter vid examination är enligt riktlinjer och regelverk vid Högskolan i Borås.

Kurslitteratur och övriga läromedel

Kurslitteraturen är huvudsakligen på engelska, men artiklar på svenska kan tillkomma.

Agrawal, R. & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. I VLDB '94, Proceedings of the 20th International Conference on Very Large Data Bases, ss. 487-499.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, *Information Fusion*, 58, ss. 82-115.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, ss. 321-357.

Drummond, C. (2009). Replicability is not Reproducibility: Nor is it Good Science. I Proceedings of the Evaluation Methods for Machine Learning Workshop at the 26th ICML.

Freitas, A. A. (2014). Comprehensible classification models: a position paper. *ACM SIGKDD Exploration Newsletter*, 15(1), ss. 1–10.

Géron, A. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems*. Sebastopol, CA: O'Reilly Media, Inc. (ca 250 sidor obligatoriskt)

Han, J., Pei, J., & Yin, Y. (2000). Mining Frequent Patterns Without Candidate Generation. I SIGMOD '00, Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, ss. 1–12

Krogh, A., & Vedelsby, J. (1995). Validation, and Active Learning. Advances in neural information processing systems, 7, ss. 231-238.

Shepperd, M., Guo, Y., Li, N., Arzoky, M., Capiluppi, A., Counsell, S., Destefanis, G., Swift, S., Tucker, A., & Yousefi, L. (2019). The Prevalence of Errors in Machine Learning Experiments. I IDEAL 2019, 20th International Conference on Intelligent Data Engineering and Automated Learning, ss. 102-110.

Tichy, W. F. (1998). Should Computer Scientists Experiment More?. IEEE Computer 31(5), ss. 32–40.

Zaki, M. J. (2000). Scalable algorithms for association mining. IEEE Transactions on Knowledge and Data Engineering, 12(3), ss. 372–390.

Zhu, X. & Goldberg, A. B. (2009). Introduction to semi-supervised learning. Synthesis lectures on artificial intelligence and machine learning 3(1): ss.1-130 (urval om ca 30 sidor).

Föreläsningsmaterial, vetenskapliga artiklar och tekniska rapporter tillkommer, enligt lärarens anvisningar. (max 50 sidor).

Studentinflytande och utvärdering

Kursen utvärderas i enlighet med gällande riktlinjer för kursvärderingar vid Högskolan i Borås, där studenternas synpunkter ska inhämtas. Kursutvärderingsrapporten publiceras och återkopplas till deltagande och blivande studenter i enlighet med ovan nämnda riktlinjer, och ligger till grund för framtida utveckling av kurser och utbildningsprogram. Kursansvarig lärare ansvarar för att utvärdering enligt ovan genomförs.

Övrigt

Kursen ges som en fristående kurs. Primär målgrupp är utvecklare av AI-baserade tillämpningar inom näringslivet.