

Introduktion till maskininlärning och artificiell intelligens för materialvetenskap, 7.5 hp

Introduction to machine learning and artificial intelligence for material science, 7.5 credits

Forskarutbildningskurs

6FIFMA6

Institutionen för fysik, kemi och biologi

Gäller från: Andra halvår 2025

Fastställd av

Fastställandedatum

Diarienummer

Behörighetskrav

- Fullgjort kursfordringar om minst 240 högskolepoäng (hp), varav minst 60 hp på avancerad nivå, eller på något annat sätt förvärvat motsvarande kunskaper
- Grundläggande kunskaper i matematik, inklusive statistik, linjär algebra och sannolikhetsteori.
- Viss grundläggande erfarenhet av programmering. En betydande del av kursen utnyttjar Python, varvid tidigare erfarenhet av Python-programmering är en fördel. Kursen inleds dock med en introduktion till Python och vetenskaplig programmering i Python, vilket säkerställer att alla studenter får möjlighet att tillägna sig tillräckliga grundkunskaper för att kunna följa resten av kursen

Särskild information

Kursen är utformad för doktorander (och eventuellt masterstudenter) med en bakgrund inom kemi, materialvetenskap, fysik eller likande områden, men med liten eller ingen erfarenhet av maskininläring och artificiell intelligens. Kursens mål är att ge en solid grund till ML/AI, och att utrusta studenterna med den kunskap och det självförtroende som krävs för att börja använda ML/AI-verktyg för att lösa problem relevant för deras forskning. Kursen kombinerar teori med praktiska programmeringsövningar och går igenom rutiner för att använda Python för att hantera verkliga datavetenskapliga problem med hjälp av ML/AI. Även om innehållet är brett tillämpligt, kommer illustrativa exempel, där det är relevant, att hämtas från materialvetenskap och kemi

Lärandemål

I slutet av kursen ska studenterna:

- Vara bekanta med vanliga maskininläringstekniker.
- Vara bekanta med olika typer av neurala nätverk.
- Vara bekanta med några mer avancerade koncept, såsom Bayesiansk optimering och generativa modeller.
- Vara bekanta med hur man kan använda materialdata från öppna online databaser.
- Baserat på ett givet problem, kunna göra ett välgrundat val av lämpliga tekniker/algoritmer att använda.
- Kunna skriva Python-kod för att använda vanliga maskininlärningsalgoritmer för att lösa regressions-, klassificerings- och klustringsproblem.
- Kunna använda Python för att lösa maskininlärningsproblem som är relevanta för materialvetenskaplig forskning.

Kursinnehåll

Kursen är indelad i fyra moduler.

1) Introduktion till Python

Denna korta modul ger en snabbkurs i Python och täcker hur man sätter upp en lokal utvecklingsmiljö samt de programmeringskoncept som behövs för att börja använda maskininläring. För studenter utan tidigare erfarenhet av Python ger denna modul den grundläggande kunskap som krävs för att kunna följa resten av kursen.

2) Klassisk maskininläring

Denna modul introducerar grundläggande maskininlärningsalgoritmer för regression, klassificering, klustring, och dimensionsreduktion. Algoritmer som tas upp inkluderar bland annat linjär regression, logistisk regression, SoftMax, k-nearest neighbors, supportvektormaskiner, naiv Bayes, beslutsträd, random forests, principal component analysis, K-means, och DBSCAN. Samtidigt går vi igenom datahantering, modellträning, kostnadsfunktioner, regularisering och modellevaluering. Python-moduler som NumPy, Pandas, SciPy och Scikit-learn introduceras i denna modul.

3) Neurala nätverk

Denna modul fokuserar på grunderna i neurala nätverk, deras inre komponenter, aktiveringsfunktioner, träningsprocesser, nätverksdesign, och regulariseringstekniker. Vanliga nätverksarkitekturer som feedforward-nätverk, convolutional networks och autoencoders kommer att presenteras. I denna modul introduceras Python-moduler som TensorFlow och Keras.

4) Avancerade ämnen

Den sista modulen introducerar några mer avancerade ämnen som Bayesiansk optimering, graph neural networks, explainable AI (inklusive SHAP-analys), generativa modeller, transformers och stora språkmodeller. Modulen omfattar också materialinformatik och feature engineering med fokus på material- och molekylära egenskaper. Python-moduler relaterade till materialinformatik, som RDKit och Pymatgen, introduceras i denna del.

Undervisnings- och arbetsformer

Kursen innehåller:

- 15 föreläsningar
- 4 datorlaborationer med hands-on uppgifter som implementerar koncept från föreläsningarna
- 1 obligatoriskt slutseminarium för redovisning av ett större projekt.

Examination

Examinationen består av två delar:

1. **Programmeringsuppgifter:** Studenterna ska lämna in uppgifter baserade på de programmeringsövningar som genomförts under kursen.
2. **Slutprojekt:** Studenterna väljer ett projektämne i samråd med kursansvarig, helst relaterat till deras egen forskning. Projektet bedöms genom en skriftlig rapport och en muntlig presentation.

Betygsskala

Tvågradig skala

Kurslitteratur

Huvudbok: Hands-On- Machine Learning with Scikit-Learn, Keras, and Tensorflow, second edition, by Aurélien Géron

Utöver huvudboken kommer komplimenterande material från olika källor distribueras under kursen.

Övrig information

Planering och genomförande av kursen skall utgå från kursplanens formuleringar. Kursvärdering samt analys och förslag som rör generell utveckling och förbättring av kursen återkopplas till Forsknings- och forskarutbildningsnämnden av kursansvarig lärare.

Om kursen upphör eller genomgår större förändring erbjuds normalt examination enligt denna kursplan vid totalt tre tillfällen inom/i anslutning till de två terminer som följer.