



LUNDS
UNIVERSITET

Naturvetenskapliga fakulteten

BERN02, Beräkningsvetenskap: Reproducerbar dataanalys och statistisk inläring, 7,5 högskolepoäng
Computational Science: Reproducible Data Science and Statistical Learning, 7.5 credits
Avancerad nivå / Second Cycle

Fastställande

Kursplanen är fastställd av Naturvetenskapliga fakultetens utbildningsnämnd 2022-12-14 att gälla från och med 2022-12-14, höstterminen 2023.

Allmänna uppgifter

Kursen är en obligatorisk kurs på avancerad nivå för en naturvetenskaplig masterexamen i tillämpad beräkningsvetenskap, samt en alternativ-obligatorisk kurs på avancerad nivå för en naturvetenskaplig masterexamen i beräkningsvetenskap. Kursen är även en valbar kurs på avancerad nivå för en naturvetenskaplig masterexamen i matematisk statistik.

Undervisningsspråk: Engelska

Huvudområde

Beräkningsvetenskap

Matematisk statistik

Tillämpad beräkningsvetenskap

Fördjupning

A1N, Avancerad nivå, har endast kurs/er på grundnivå som förkunskapskrav

A1N, Avancerad nivå, har endast kurs/er på grundnivå som förkunskapskrav

A1N, Avancerad nivå, har endast kurs/er på grundnivå som förkunskapskrav

Kursens mål

Kursens övergripande mål är att låta studenterna arbeta med och kombinera två områden för dataanalys inom beräkningsvetenskap: reproducerbara arbetsflöden och statistisk inläring. Detta innebär att kunna skapa rapporter där programmeringskod, resultat och rapporttext kombineras i samma dokument, tillämpat på ett urval av vanliga metoder i statistisk parametrisk modellering och maskininläring.

Kunskap och förståelse

Efter avslutad kurs ska studenten kunna:

- beskriva principer för reproducerbara och interoperabla arbetsflöden för dataanalys och rapportering av resultat
- redogöra för ett urval av metoder för statistisk modellering och algoritmer för maskininlärning
- identifiera lämpliga modeller för klassificering, parameterskattning och prediktion för en given forskningsfråga.

Färdighet och förmåga

Efter avslutad kurs ska studenten kunna:

- skapa, dela och använda verktyg för "literate programming" t.ex. Jupiter Notebooks, analytiska arbetsflöden och versionshantering för dataanalys och rapportering
- hantera, presentera och visualisera data för att lyfta fram viktiga egenskaper i ett datamaterial
- tillämpa ett urval av vanliga metoder för statistisk parametrisk modellering och maskininlärning
- använda ett programmeringsspråk, t.ex. Python eller R, för statistisk analys och maskininlärning
- redovisa metod, resultat och slutsatser från en dataanalys i en reproducerbar och tolkningsbar skriftlig rapport.

Värderingsförmåga och förhållningssätt

Efter avslutad kurs ska studenten kunna:

- resonera kring nytta och begränsningar för modeller och metoder valda för ett praktiskt problem
- kritiskt granska reproducerbarhet och interoperabilitet i arbetsflöden för att hantera och analysera vetenskapliga data.

Kursens innehåll

Kursen introducerar grundläggande principer för reproducerbara och interoperabla arbetsflöden med ett tydligt fokus på tillämpning. Studenter får en översikt i import, transformering och visualisering av data, där verklighetstroga data förbereds för analys i elektroniska "notebooks". Dessa elektroniska "notebooks" använder verktyg för "literate programming", analytiska arbetsflöden och versionshantering.

Studenter introduceras till ett urval av metoder för statistisk inlärning. Dessa inkluderar generaliserad lineär regression, med maximum likelihood och Bayesiansk inferens för att skatta parametrar, maskininlärningsmetoder för regression och klassificering, och metoder för dimensionsreducering och klustering. Kursen tar också upp generella metoder för modellutvärdering (t.ex. korsvalidering) och modellval. Examinationen baseras på projekt där studenterna ska välja lämpliga metoder för analys av givna data, och utföra analysen i ett reproducerbart och interoperabelt arbetsflöde, som sammanfattas i en rapport.

Kursens genomförande

Undervisningen utgörs av föreläsningar, datorlaborationer och projektarbeten. Deltagande i datorlaborationer och projektarbeten är obligatoriskt.

Kursens examination

Examination sker i form av projektarbeten under kursens gång som inkluderar skriftliga projektrapporter, en muntlig presentation och kamratgranskning av ett annat studentprojekt, samt genom laborationer och en muntlig tentamen vid kursens slut.

För studerande som ej godkänts vid ordinarie tentamen erbjuds ytterligare tentamenstillfälle i nära anslutning härtill.

Om så krävs för att en student med varaktig funktionsnedsättning ska ges ett likvärdigt examinationsalternativ jämfört med en student utan funktionsnedsättning, så kan examinator efter samråd med universitetets avdelning för pedagogiskt stöd fatta beslut om alternativ examinationsform för berörd student.

Prov/moment för denna kurs finns i en bilaga i slutet av dokumentet.

Betyg

Betygsskalan omfattar betygsgraderna Underkänd, Godkänd, Väl godkänd. För att bli godkänd på hela kursen krävs godkända projektrapporter, godkänd muntlig presentation och kamratgranskning, godkänd muntlig tentamen samt godkända laborationer med tillhörande obligatoriska moment.

Betygsskalan för projektrapporter, muntlig presentation, kamratgranskning och laborationer är Underkänd, Godkänd, medan den muntliga tentamen betygsätts enligt betygsskala Underkänd, Godkänd, Väl godkänd.

Slutbetyget avgörs genom betyg på muntlig tentamen.

Förkunskapskrav

För tillträde till kursen krävs 90 hp i naturvetenskapliga eller tekniska studier, inklusive 43,5 hp i matematik, varav 7,5 hp i statistik och 6 hp i programmering, samt Engelska 6/B.

alternativt

en kandidatexamen i fysik samt Engelska 6/B.

Övrigt

Kunskaper inom Python eller R rekommenderas, men är inte ett krav.

Prov/moment för kursen BERN02, Beräkningsvetenskap: Reproducerbar dataanalys och statistisk inläring

Gäller från H23

- 2301 Muntlig tentamen, 3,0 hp
Betygsskala: Underkänd, Godkänd, Väl godkänd
- 2302 Projektarbete -rapporter, presentation och kamratgranskning, 4,0 hp
Betygsskala: Underkänd, Godkänd
- 2303 Laborationer, 0,5 hp
Betygsskala: Underkänd, Godkänd