



LUNDS UNIVERSITET
Lunds Tekniska Högskola

Kursplan för

Dataanalys: statistisk inlärning och visualisering med projekt

Data Analysis: Statistical Learning and Visualization with project

FMSF90, 7,5 högskolepoäng, G2 (Grundnivå, fördjupad)

Gäller för: Läsåret 2023/24

Fakultet: Lunds tekniska högskola

Beslutad av: Programledning I

Beslutsdatum: 2023-04-14

Allmänna uppgifter

Huvudområde: Teknik.

Valfri för: F4, Pi4, R4

Undervisningsspråk: Kursen ges på engelska

Syfte

Kursen inleds med en överblick av grundläggande datahantering och visualisering, med fokus på att kunna identifiera och illustrera olika egenskaper och särdrag hos data.

Därefter presenteras viktiga metoderna inom modern statistisk inlärning. Stor vikt läggs vid dimensionsreduktion, övervakade och oövervakade inlärning. Problemen med att anpassa flera olika modeller (multiple-testing) och metodernas relation till regression diskuteras. Datorbaserade laborationer och projekt utgör en viktig lärandeaktivitet. Kursen avslutas med ett projekt där studenterna ska välja lämpliga metoder för att analysera ett givet data material.

Mål

Kunskap och förståelse

För godkänd kurs skall studenten

- Beskriva olika sätt att aggregera, sammanfatta och visualisera data.
- Förklara principerna för dimensions reduktion

- Förklara principerna några olika typer övervakad och oövervakad inlärning.

Färdighet och förmåga

För godkänd kurs skall studenten

- kunna hantera, presentera och visualisera data för att belysa viktiga egenskaper och kännetecken i ett komplext datamaterial.
- kunna utföra dimensions reduktion och imputation av saknad data.
- kunna använda vanliga metoder för klassificering, övervakad inlärning och oövervakad inlärning
- kunna dra slutsatser om data baserat på resultaten från metoder för klassificering och inlärning.
- redovisa analys av och slutsatser från ett praktiskt problem i en skriftlig rapport.

Värderingsförmåga och förhållningssätt

För godkänd kurs skall studenten

- Reflektera över den valda modellens och metodens begränsningar samt möjliga alternativa lösningsmetoder.
- Reflektera över eventuella problem med att anpassa flera olika modeller till samma datamaterial.

Kursinnehåll

- Grundläggande datahantering och vanliga visualisering metoder för data.
- Metoder för data reduktion som PrincipalKomponentAnalys (PCA) och deras användning för imputation av saknad data.
- Metoder för oövervakad och övervakad inlärning/klassificering som: StödVektorMaskin (SVM), klustring (K-means), hierarkisk klustring, enklare regressionsmetoder samt metoder med beslutsträd (bagging, boosting och random forests).
- Multiple-testning och lösningar som Benjamini-Hochberg och Bonferroni.

Kursens examination

Betygsskala: TH - (U,3,4,5) - (Underkänd, Tre, Fyra, Fem)

Prestationsbedömning: Slutbetyget ges av det avslutande projektet.

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.

Delmoment

Kod: 0123. **Benämning:** Datorlaboration 1.

Antal högskolepoäng: 2. **Betygsskala:** UG. **Prestationsbedömning:** Skriftlig rapport. **Delmomentet omfattar:** Datahantering and visualisering.

Kod: 0223. **Benämning:** Datorlaboration 2.

Antal högskolepoäng: 2. **Betygsskala:** UG. **Prestationsbedömning:** Skriftlig rapport. **Delmomentet omfattar:** Övervakad inläring.

Kod: 0323. **Benämning:** Datorlaboration 3.

Antal högskolepoäng: 2. **Betygsskala:** UG. **Prestationsbedömning:** Skriftlig rapport. **Delmomentet omfattar:** Övervakad inläring.

Kod: 0423. **Benämning:** Projekt.

Antal högskolepoäng: 1,5. **Betygsskala:** TH. **Prestationsbedömning:** Skriftlig projektrapport **Delmomentet omfattar:** Avslutande projekt

Antagningsuppgifter

Förkunskapskrav:

- FMAB30 Flerdimensionell analys eller FMAB35 Flerdimensionell analys med vektoranalys eller FMSF20 Matematisk statistik, allmän kurs eller FMSF25 Matematisk statistik - kompletterande projekt eller FMSF32 Matematisk statistik eller FMSF45 Matematisk statistik, allmän kurs eller FMSF50 Matematisk statistik, allmän kurs eller FMSF55 Matematisk statistik, allmän kurs eller FMSF70 Matematisk statistik eller FMSF75 Matematisk statistik, allmän kurs eller FMSF80 Matematisk statistik, allmän kurs
- FMAA20 Linjär algebra med introduktion till datorhjälpmedel eller FMAA21 Linjär algebra med numeriska tillämpningar eller FMAB20 Linjär algebra eller FMSF20 Matematisk statistik, allmän kurs eller FMSF25 Matematisk statistik - kompletterande projekt eller FMSF32 Matematisk statistik eller FMSF45 Matematisk statistik, allmän kurs eller FMSF50 Matematisk statistik, allmän kurs eller FMSF55 Matematisk statistik, allmän kurs eller FMSF70 Matematisk statistik eller FMSF75 Matematisk statistik, allmän kurs eller FMSF80 Matematisk statistik, allmän kurs

Förutsatta förkunskaper: En grundkurs i matematisk statistik samt kunskaper i linjär algebra

Begränsat antal platser: 50

Urvalskriterier: Avklarade högskolepoäng inom programmet. (Observera att endast högskolepoäng som enligt Ladok tillgodoräknats inom programmet före antagningen räknas. För studenter på masterprogram adderas 180 hp motsvarande tidigare kandidatexamen.) Förtur ges till studenter vars program har kursen listad i läro- och timplanen. Bland dessa studenter ges platsgaranti till studerande på specialiseringen i Riskmodellering inom civilingenjörsprogrammet i Risk, säkerhet och krishantering.

Kursen överlappar följande kurser: FMSF86, FMAN45, EDAN96

Kurslitteratur

- Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani: An Introduction to Statistical Learning, with Applications in R (2ed, 2021 Edition). Springer, 2021, ISBN: 978-1071614174. Finns som e-bok:
https://web.stanford.edu/~hastie/ISLRv2_website.pdf.
- Jake VanderPlas: Python Data Science Handbook, Essential Tools for Working with Data. O'Reilly, 2016, ISBN: 978-1491912058. Finns som e-bok.

Kontaktinfo och övrigt

Studierektor: Johan Lindström, studierektor@matstat.lu.se

Kursansvarig: Linda Hartman, linda.hartman@matstat.lu.se

Hemsida: <https://www.maths.lu.se/utbildning/civilingenjoersutbildning/matematisk-statistik-paa-civilingenjoersprogram/>

Övrig information: Ges tillsammans med FMSF86. Endast en av kurserna FMSF86 och FMSF90 får ingå i examen. Kursen överlappar EDAN96.