



SF1930 Statistisk inlärning och dataanalys 6,0 hp

Statistical Learning and Data Analysis

När kurs inte längre ges har student möjlighet att examineras under ytterligare två läsår.

Fastställande

Skolchef vid SCI-skolan har 2020-10-07 beslutat att fastställa denna kursplan att gälla från och med HT 2021, diarienummer: S-2020-1415.

Betygsskala

A, B, C, D, E, FX, F

Utbildningsnivå

Grundnivå

Huvudområden

Teknik

Särskild behörighet

Slutförd grundkurs i sannolikhets teori och statistik (SF1918, SF1922 eller motsvarande).

Undervisningsspråk

Undervisningsspråk anges i kurstillfällesinformationen i kurs- och programkatalogen.

Lärandemål

Efter avslutad kurs ska studenten kunna:

- formulera och tillämpa begrepp inom statistisk inferens och prediktion för att lösa teoretiska problem;
- formulera och tillämpa metoder inom statistisk inferens och prediktion för att lösa problem inom dataanalys;
- utforma och implementera metoder inom statistisk inlärning för dataanalys.

Kursinnehåll

Kursen ger en introduktion till teorin för statistisk inferens och prediktion, vilka utgör de grundläggande målen för modern statistisk dataanalys och maskininlärning. Stor vikt ges åt flerdimensionella sannolikhetsfördelningar och exponentialfamiljer, vilka är fundamentala verktyg för att modellera dataanalytiska problem, och med hjälp av grafiska modeller erhålls en kraftfull teori för att beskriva betingade beroenden med bäring på statistiska inferensproblem av hög dimension. Beslutsteori ger ett ramverk för att dels fatta optimala beslut under statistisk osäkerhet, dels väga olika statistiska tillvägagångssätt mot varandra. Speciellt spelar Bayesiansk beslutsteori – i vilken inferens- och inlärningsproblemen löses genom beräkning av posteriorfördelningar respektive prediktiva fördelningar – idag en central roll inom statistisk dataanalys och används för att konstruera Bayesianska punktskattningar, hypotestest och kredibilitetsintervall. Parallellt med det Bayesianska synsättet diskuteras även likelihood-teori och i synnerhet maximum likelihood-skattningens asymptotiska egenskaper då mängden data växer mot oändligheten. I kursen introduceras även grundläggande statistiska beräkningsmetoder, såsom stokastiska gradientmetoder och Markovkedjemetoder (MCMC). Sådana metoder är av stor betydelse inom modern datorintensiv statistik och tillämpas i kursen på verkliga dataanalytiska problem inom ramen för ett datorbaserat projekt.

Examination

- INL1 - Inlämningsuppgift, 2,0 hp, betygsskala: P, F
- TEN1 - Skriftlig tentamen, 4,0 hp, betygsskala: A, B, C, D, E, FX, F

Examinator beslutar, baserat på rekommendation från KTH:s handläggare av stöd till studenter med funktionsnedsättning, om eventuell anpassad examination för studenter med dokumenterad, varaktig funktionsnedsättning.

Examinator får medge annan examinationsform vid omexamination av enstaka studenter.

Etiskt förhållningssätt

- Vid grupparbete har alla i gruppen ansvar för gruppens arbete.
- Vid examination ska varje student ärligt redovisa hjälp som erhållits och källor som använts.
- Vid muntlig examination ska varje student kunna redogöra för hela uppgiften och hela lösningen.

