

Data Mining and Statistical Learning

Fristående- och programkurs

15 hp

Data Mining and Statistical Learning

732A33

Gäller från:

Fastställd av

Filosofiska fakultetens kvalitetsnämnd

Fastställandedatum

2008-09-10

Revideringsdatum

2013-03-18

Huvudområde

Statistik

Utbildningsnivå

Avancerad nivå

Fördjupningsnivå

A1X

Förkunskapskrav

För tillträde till kursen krävs det att den studerande har ett kandidatexamen med lägst 90 hp, dvs. 18 månader heltidsstudier, i matematik, tillämpad matematik, statistik eller datavetenskap. Matematikkurserna på grundnivå bör inkludera såväl kalkyl som linjär algebra. Utöver detta, erfordras kurser på grundnivå i grundläggande statistik och datavetenskap. Engelska B eller motsvarande.

Lärandemål

Kursen lägger grunden för ett professionellt arbete och forskning där stora datavolymer undersöks, modifieras, modelleras och utvärderas för att upptäcka tidigare okända mallar och trender.

Efter avslutad kurs skall den studerande kunna:

- redovisa principer av statistisk modellering, speciellt för analyser av stora datamängder
- använda redskap i SAS miljö för att utforska stora och komplexa datamängder, skapa datadrivna modeller, utvärdera deras utfall och använda sådana modeller för prediktion.
- jämföra prestanda av statistiska och data mining modeller för att välja den mest relevanta modellen i ett givet sammanhang.

Kursinnehåll

- grundläggande begrepp i statistical learning, speciellt inom övervakad inlärning
- modellvalsstrategier som använder tränings- validerings- och testmängder och modellvalet med hjälp av korsvalidering
- linjär regression och krympningsmetoder
- splines och kernelmetoder
- beslutsträd och klassificeringsmetoder såsom diskriminantanalys och logistisk regression
- neurala nätverk, supportvektormaskiner och generaliserade additiva modeller
- samspelsmetoder, inklusive bagging och boosting
- Bayesianiska metoder i data mining

Undervisnings- och arbetsformer

Kursen består av föreläsningar, datorövningar och seminarier. Föreläsningarna ägnas åt genomgång av teori, koncept och metodik. Datorövningarna ägnas åt praktisk dataanalys i SAS miljö (som regel) eller i andra miljöer (i undantagsfall). Seminarier ägnas åt studentpresentationer och diskussioner av uppgifter.
Språk: Engelska

Examination

Skriftlig redovisning av labbuppgifter. Obligatorisk närvaro på seminarierna. En skriftlig tentamen.

Betygsskala

ECTS, EC

Övrig information

Planering och genomförande av kurs skall utgå från kursplanens formuleringar. Den kursvärdering som skall ingå i varje kurs skall därför behandla frågan om hur kursen överensstämmer med kursplanen. Kursen bedrivs på ett sådant sätt att både mäns och kvinnors erfarenhet och kunskaper synliggörs och utvecklas.

Institution

Institutionen för datavetenskap