



INSTITUTIONEN FÖR MATEMATISKA VETENSKAPER

MSA220 Statistisk slutledning för stora datamängder, 7,5 högskolepoäng

Statistical Learning for Big Data, 7.5 credits

Avancerad nivå / Second Cycle

Fastställande

Kursplanen är fastställd av Institutionen för Matematiska vetenskaper 2015-02-27 och senast reviderad 2019-05-14. Den reviderade kursplanen gäller från och med 2019-09-02, höstterminen 2019.

Utbildningsområde: Naturvetenskapligt 100 %

Ansvarig institution: Institutionen för matematiska vetenskaper

Inplacering

Kursen kan ingå i följande program: 1) Matematiska vetenskaper, masterprogram (N2MAT) och 2) Applied Data Science masterprogram (N2ADS)

Huvudområde

Matematisk statistik

Fördjupning

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

Förkunskapskrav

En grundläggande kurs i statistisk slutledning samt kursen *MSG500 Linjära statistiska modeller*.

Lärandemål

Efter godkänd kurs ska studenten kunna:

- visa förståelse för centrala begrepp och idéer rörande klassificering, klustring och dimensionsreducering
- lösa högdimensionella dataanalys-övningar och tolka resultaten av sådana analyser.

Innehåll

- Överblick över högdimensionell dataanalys
- Klassificering: Bayes formel, diskriminantanalys-metoder, närmaste granne klassificering, klassificering- och regressions-träd.
- Målfunktioner, optimering och gradient descent, kors-validering.
- Logistisk regression
- Regulariseringsmetoder. Gles logistisk regression, gles diskriminantanalys.
- Ensemble-metoder: bagging, random projections, random forests.
- Klustring: k-means, hierarkisk klustring, modell-baserad klustring, spektrala metoder.
- Dimensionsreduktion: PCA, kanonisk korrelation, multi-dimensional scaling.
- Speciella teman (urval av följande): nätverk och grafiska modeller, gles kovariansestimering, klustring av nätverk och community detection, neurala nätverk, matriskomplettering, collaborative filtering.
- Stor-skaliga metoder: stochastic searches, batch-metoder, online learning.

Former för undervisning

Föreläsningar, diskussioner och läsuppgifter.

Undervisningsspråk: engelska

Former för bedömning

Muntlig och/eller skriftlig examination.

Om student som underkänts två gånger på samma examinerande moment önskar byte av examinator inför nästa examinationstillfälle, ska sådan begäran inlämnas skriftligt till kursansvarig institution och bifallas om det inte finns särskilda skäl däremot (HF 6 kap § 22).

I det fall en kurs har upphört eller genomgått större förändringar ska studenten i normalfallet garanteras tillgång till minst tre provtillfällen (inklusive ordinarie provtillfälle) under en tid av åtminstone ett år med utgångspunkt i kursens tidigare uppläggning.

Betyg

På kursen ges något av betygen Väl godkänd (VG), Godkänd (G) och Underkänd (U).

Kursvärdering

Resultatet och eventuella förändringar i kursens upplägg ska förmedlas både till de studenter som genomförde värderingen och till de studenter som ska påbörja kursen.