Introduktion till maskininlärning

Fristående- och programkurs

9 hp

Introduction to Machine Learning

732A95

Gäller från: 2016 HT

Fastställd av
Filosofiska fakultetens kvalitetsnämnd

Fastställandedatum
2016-04-01

Gavs sista gången
HT 2017

Huvudområde
Statistik

Utbildningsnivå
Avancerad nivå

Fördjupningsnivå
A1X

Kursen ges för
- Master´s Programme in Statistics and Data Mining

Förkunskapskrav
Kandidatexamen i något av följande ämnen: statistik, matematik, tillämpad matematik, datavetenskap, teknik eller motsvarande examen. Utöver detta, erfordras godkända/avklarade kurser i kalkyl, linjär algebra.
Studenten ska också ha följande kurser godkända: En grundkurs i statistik om minst 6 hp; en kurs i programmering om minst 6 hp.
Engelska B eller motsvarande.

Lärandemål
Efter avslutad kurs skall den studerande på en avancerad nivå kunna:

- använda relevanta begrepp och metoder inom maskininlärning för att formulera, strukturera och lösa praktiska problem som har anknytning till stora eller komplexa datamängder,
- göra inferens för parametrarna i ett antal vanligt förekommande maskininlärningsmodeller,
- använda maskininlärningsmetoder för prediktion och beslutsstöd,
- utvärdera kvalitén av maskininlärningsmodeller,
- välja en lämplig modell i situationer med begränsad eller saknad information om bakomliggande samband i datamaterialet,
- implementera maskininlärningsmodeller i ett programmeringsspråk och använda existerande maskininlärningsmjukvaror för att analysera stora eller komplexa datamängder, göra prediktioner och utvärdera osäkerhet av dessa prediktioner.

Kursinnehåll

Kursen fokuserar på huvudbegrepp och huvudredskap i probabilistisk maskininlärning som är nödvändiga för ett professionellt arbete och forskning inom dataanalys.

- introduktion och översikt av maskininlärning (inklusive regression, klassificering, övervakad och oövervakad inlärning) och dess tillämpningsområden,
- Nearest Neighbors and Naïve Bayes,
- diskriminantanalys, logistisk regression och beslutsträd,
- modellval och osäkerhetsskattning: holdout metoden, korsvalidering, AIC, bootstrap och konfidensintervall,
- linjär regression och regulariseringsmetoder (Ridge, LASSO),
- splines, generaliserade linjära och additiva modeller,
- principalkomponentanalys (PCA) och principalkomponentregression (PCR),
- kernel utjämnare, kernel trick och supportvektormaskiner,
- neurala nätverk,
- bagging, boosting och random forests,
- Online learning och mixture models.

Undervisnings- och arbetsformer

Kursen består av föreläsningar, datorövningar och seminarier som kompletteras med självstudier. Föreläsningarna ägnas åt genomgång av teori, koncept och metodik. Datorövningarna ägnas åt praktisk dataanalys i en med hjälp av maskininlärningsmjukvara. Seminarier ägnas åt studentpresentationer och diskussioner av uppgifter.
Undervisningsspråk: engelska.

Examination

Skriftlig redovisning av labbuppgifter. Obligatorisk närvaro på seminarierna. En skriftlig tentamen. Detaljerad information återfinns i studiehandledningen.

Betygsskala

Tregradig skala, U, G, VG

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