Masters Programme in Statistics and Machine Learning

120 hp

Masters Programme in Statistics and Machine Learning

F7MSL

Gäller från: 2018 HT

Fastställd av
Fakultetsstyrelsen för filosofiska fakulteten

Fastställandedatum
2017-11-24

Inledning

Som en följd av en snabb IT utveckling har samhället blivit översvämmad av enorma volymer av information som genereras av stora eller komplexa system. Informationen kan largas i stora databaser, inkomma som en ström eller vara resultat av interaktion mellan systemet och lärande miljön. Detta program syftar på inlärning av dessa komplexa informationsvolymer till modeller och algoritmer som förbättrar analyser, prognoser och beslutsfattandet. Statistisk modellering och analys är integrerade med maskininlärning, data mining och datahantering för att forma en grundlig bas för ett professionellt arbete med informationsmodellering och dataanalys i stora eller komplexa system. Programmet skapar också utmärkta förutsättningar för att göra en forskningskarriär.

Mål

Kunskap och förståelse

Efter genomgången utbildning ska den studerande

- visa kunskap och förståelse inom huvudområdet statistik, inbegripet såväl brett kunnande inom området som väsentligt fördjupade kunskaper inom dess bransch, maskininlärning, samt fördjupad insikt i aktuellt forsknings- och utvecklingsarbete, och
- visa fördjupad metodkunskap inom statistik.

Fördjupade kunskaper inom maskininlärning skall inkludera moderna kraftiga metoder för klassificering och regression, prediktion, metoder för statistisk simulering och optimering, Bayesianska metoder och analys av stora databaser..

Färdighet och förmåga

Efter genomgången utbildning ska den studerande

- visa förmåga att kritiskt och systematiskt integrera kunskap och att analysera, bedöma och hantera komplexa företeelser, frågeställningar och situationer även med begränsad information,
- visa förmåga att kritiskt, självständigt och kreativt identifiera och formulera frågeställningar, att planera och med adekvata metoder genomföra kvalificerade uppgifter inom givna tidsramar och därigenom bidra till kunskapsutvecklingen samt att utvärdera detta arbete,
- visa förmåga att i såväl nationella som internationella sammanhang muntligt och skriftligt klart redogöra för och diskutera sina slutsatser och den kunskap och de argument som ligger till grund för dessa i dialog med olika grupper, och
- visa sådan färdighet som fordras för att delta i forsknings- och utvecklingsarbete eller för att självständigt arbeta i annan kvalificerad verksamhet.

Värderingsförmåga och förhållningssätt

Efter genomgången utbildning ska den studerande

- visa förmåga att inom huvudområdet statistik göra bedömningar med hänsyn till relevanta vetenskapliga, samhälleliga och etiska aspekter samt visa medvetenhet om etiska aspekter på forsknings- och utvecklingsarbete,
- visa insikt om vetenskapens möjligheter och begränsningar, och speciellt möjligheter och begränsningar av statistik, dess roll i samhället och människors ansvar för hur den används, och
- visa förmåga att identifiera sitt behov av ytterligare kunskap och att ta ansvar för sin kunskapsutveckling.

Efter genomgången utbildning ska den studerande kunna

- modellera informationsvolymer som generas av stora eller komplexa system
- välja en lämplig analysmodell i ett givet sammanhang
- utvinna och strukturera stora och komplexa datamängder
- utforska, sammanfatta och presentera stora och komplexa datamaterial med hjälp av statiska, interaktiva och dynamiska visualiseringsredskap
- använda avancerade mjukvaror för att analysera stora eller komplexa datavolymer
- implementera modeller för dataanalys, prediktion och beslutsfattande i något programmeringsspråk
- kombinera datamaterialet och andra källor av prioriinformation för att förbättra inferensen och prediktionsförmågan
- ge exempel på tillämpningsområden där det behövs analys och modellering av informationsvolymer som kommer från stora eller komplexa system
- upptäcka och statistiskt granska tidigare okända mallarna och trenderna i en datamängdpresentera en uppsats som innehåller en teoretisk eller en tillämpad studie av stora eller komplexa system eller datamängder med hjälp av metoder från statistik och maskininlärning.

Innehåll

Programmet kombinerar kurser i statistik, datavetenskap och matematik. Obligatoriska kurser, inledande kurser och en masteruppsats på 30 hp främjar progression och djupet av förståelsen. Inledande kurser erbjuds för att fylla i brister i studenternas kunskaper och för att se till att studenterna är ordentligt förberedda till programmets kurser.

Obligatoriska kurser

Akademiska studier på avancerad nivå, 3 hp (ges termin 1)

Målet för denna kurs är att förbereda studenterna för akademiska studier på avancerad nivå, samt att lära ut ett akademiskt förhållningssätt i stort. En

grundläggande ambition är att tillhandahålla väsentliga redskap för att studera på avancerad nivå i Sverige. Dessutom kommer programspecifika moment diskuteras.

Maskininlärning (ges termin 1)
Grundläggande koncept inom maskininlärning och data mining. Bayesiansk och frekventastes modellering, modelval. Linjär regression och regularisering. Linjär diskriminantanalys och logistisk regression. Bagging och boosting. Splines, generaliserade additiva modeller, beslutsträd och random forest. Kernel utjämning och stödvektormaskiner. Gaussiansk process.

Data Mining, 6 hp (ges termin 2)
Statistiska principer och redskap för uppdelning av objekt i grupper och utvinning av samband som är gömda i stora datamängder. Partitionell och hierarkisk klustring. Klustervärdering. Associationsanalys med hjälp av enhetsmängder och associationsregler. Utvärdering av associationsregler.

Analys av Big Data, 6hp (ges termin 2)
Filsystem och databaser för Big Data. Förfrågningar till Big Data. Resurshantering i en klustermiljö. Parallerisering av beräkningar för Big Data. Maskininlärning för Big Data.

Introduktion till Python, 3 hp (ges termin 2)
Python programmeringsmiljö. Datastrukturer:tal, strängar, listor, tuples, lexikon. Grundläggande språkelement: loopar, villkor, funktioner. Moduler. Input och output. Debuggning. Maskininlärning och datamining i Python.

Vetenskapsfilosofi, 3 hp (ges termin 2)
Naturlagarna och vetenskapliga modeller. Samband mellan teorier och observationer. Krafter som påverkar vetenskapliga förändringar.

Bayesianska metoder, 6 hp (ges termin 2)
Bayes sats för att kombinera datamängder med prioriinformation. Bayesiansk analys av konjugerade modeller. Markov Kedjor Monte Karlo för Bayesianska beräkningar. Bayesianska modellvalet.

Datorintensiva statistiska metoder, 6 hp (ges termin 2)
Datorernas aritmetik. Slumptalgenererings- och simuleringsmetoder. Markov Kedjor Monte Karlo. Numerisk linjär algebra. Optimeringsmetoder i statistik.

Profilkurser

Visualisering, 6 hp (ges termin 1 för studenter antagna jämna år och termin 3 för studenter antagna ojämna år)
Avancerade visualiseringsmetoder för stora och komplexa datamängder. Interaktiva och dynamiska statistiska diagram. Visualisering av spatial information.

Avancerad maskininlärning, 6 hp (ges termin 3)

Bayesianska nätverk och dolda Markovmodeller. State-space modeller och slumpfält. Neurala nätverk. Principer av djupinlärning och dess redskap: djupa neurala nätverk, Boltzman maskiner.

Tidsserieanalys, 6 hp (ges termin 1 för studenter antagna ojämna år och termin 3 för studenter antagna jämna år)
Tidsseriedekomposition. Autokorrelation och partiell autokorrelation. Prognoser med hjälp av regression av tidsserier, ARIMA modeller och transferfunktioner. Interventionsanalys. Trendutvinning.

Multivariata Statistiska Metoder, 6 hp (ges termin 1 för studenter antagna ojämna år och termin 3 för studenter antagna jämna år)
Analys av korrelation- och kovariansstrukturer, inklusive principalkomponennter, faktoranalys, och kanonisk korrelation. Klassificering- och diskrimineringsmetoder. Flerdimensionell inferens.

Sannolikhetslära, 6 hp (ges termin 3)
Flerdimensionella slumpvariabler och betingade sannolikheter. Fördelningar av största och minsta värden i ett stickprov. Karaktäristiska funktioner och transformer. Multivariat normalfördelning. Sannolikhetsrelaterade konvergensbegrepp.

Beslutsteori, 6 hp (ges termin 3)
Resonemang med sannolikheter och likelihood-teori. Bayesiansk hypotesprövning. Beslutsteoretiska element. Nyttighets- och förlustfunktioner. Grafisk modellering som ett instrument för beslutsfattandet. Sekventiell analys.

Kompletterande kurser

Webbprogrammering 6 hp (ges termin 2)
Översikt av WWW, HTML, etc. I denna del ges en översikt av WWW, Internet, browsers, HTML, client-servers. Tekniker som Python, Flask, SQL, Websockets, JSON och andra server-side tekniker.

Bioinformatik 6 hp (ges termin 1 för studenter antagna jämna år och termin 3 för studenter antagna ojämna år)
Grunder av molekylär biologi och genetik. Gömda Markov kedjor, genetisk sekvensanalys, sekvenslikhet, sekvensgruppering. Återuppbyggnad av fylogeni, kvantitativ dragmodellering. Analys av mikromatriser. Nätverksbiologi.

Neurala nätverk och inlärningssystem, 6 hp (ges termin 2)
Oövervakad inlärning: principalkomponentanalys, analys avoberoende komponenter, vektorkvantifiering. Övervakad inlärning: neurala nätverk, radiala basfunktioner, support vector machines. Förstärkningslära: Markovprocesser, Q-learning, genetiska algoritmer.

Forskingsprojekt, 6 hp (ges termin 3)
En projektkurs där den studerande utvecklar, förbättrar eller jämför modeller eller algoritmer för maskininlärning eller data mining för ett valt

forskningsproblem.

Textmining, 6 hp (ges termin 3)
Utvinning av textinformation från olika källor. Textbearbetning med hjälp av beräkningslingvistiska metoder. Statistiska modeller för textklassificering och textprediktion.

Databasmetoder, 6 hp (ges termin 3)
En databashanterare (DBMS). Metoder för datamodellering och databasdesign. ER-diagram, relationsdatabaser och datasturkturer för databaser. Dataarkitektur och urvalsspråk för relationsdatabaser. Relationsalgebra and urvalsoptimering.

Inledande kurser

Statistiska metoder, 6 hp (ges termin 1)
Sannolikhetsbegreppet. Slumpvariabel, vanliga statistiska fördelningar och deras egenskaper. Punkt- och intervallskattning. Hypotesprövning. Enkel och multipel linjär regression. Sampling.

Avancerad R programmering, 6 hp (ges termin 1)
R miljö. Grundläggande programmeringsmetoder. Språkelement i R: variabler, vektorer, dataramar. Redskap i R: operatorer, loopar, villkor, funktioner. Importing data från text och webben. Debuggning, parallell programmering och prestandaförbättringsredskap. Statistiska och datamining redskap i R. Grafiska funktioner.Objektorienterad programmering. Effektivitetsförbättring och parallell programmering.

Masteruppsats, 30 hp
En teoretisk eller en tillämpad studie av en komplex datamängd med hjälp av statistiska och dataminingsmetoder.

Undervisnings- och arbetsformer

Programmets kurser består av föreläsningar, datorlaborationer och seminarier. Föreläsningarna ägnas åt teorier, begrepp och metoder. Datorlaborationer ger praktisk erfarenhet av dataanalys och andra metoder. Seminarier ägnas åt studentpresentationer och diskussioner av uppgifter. Kurserna vars namn innehåller ”projekt” har endast handledning.

Examination

Kurser som omfattar minst 4,5 hp examineras genom en eller flera uppgifter och en skriftlig examination. Projektkurser och masteruppsatsen examineras genom en skriftlig rapport och genom muntligt försvar av densamma.

Betyg

Betyg på kurs anges i respektive kursplan.

Förkunskapskrav

Kandidatexamen i något av följande ämnen: statistik, matematik, tillämpad matematik, datavetenskap, teknik eller motsvarande examen. Utöver detta, erfordras kurser i matematisk analys, linjär algebra, statistik och programmering.

Engelska B/Engelska 6 eller motsvarande.

Tillträdeskrav till högre termin eller kurser

För att bli behörig till termin 2 skall den studerande ha uppnått minst 6 hp på de kurser som ingår i termin 1.

För att bli behörig till termin 3 skall den studerande ha uppnått minst 40 hp på de kurser som ingår i termin 1 och 2

För att bli behörig till termin 4 skall den studerande ha uppnått minst 65 hp på de kurser som ingår i termin 1, 2 och 3, inklusive alla programmets obligatoriska kurser.

Examenskrav

En student inom programmet kan erhålla ett examensbevis med beteckningen Filosofie masterexamen med huvudområdet Statistik givet att studenten har avslutat kurser motsvarande 90 högskolepoäng som inkluderar obligatoriska kurser motsvarande 42 högskolepoäng, inledande kurser motsvarande minst 6 högskolepoäng, profilkurser motsvarande minst 12 högskolepoäng och eventuellt några kompletterande kurser. Studenten skall ytterligare ha avslutat den obligatoriska masteruppsatskursen som omfattar 30 högskolepoäng.

Examensbevis utfärdas av Filosofiska fakultetsstyrelsen, efter begäran av den studerande.

Examensbenämning på svenska

Filosofie masterexamen i huvudområdet statistik

Examensbenämning på engelska

Master of Science (120 Credits) with a major in Statistics

Särskild information

Tillgodoräknande

Filosofiska fakultetsstyrelsen eller person som utsetts av styrelsen beslutar huruvida tidigare utbildning kan överföras till programmet.

Antagningsförfarande

Studerande antas till programmet i dess helhet.

Programplan

Termin 1 (HT 2018)

Kurskod	Kursnamn	Hp	Nivå	Veckor	Block	VOF
732A60	Akademiska studier på avancerad nivå	3	A1X	v201834-201842	-	O
732A76	Forskningsprojekt	6	A1N	v201834-201843	-	V
732A93	Statistical Methods	6	A1X	v201835-201842	-	V
732A94	Avancerad programmering i R	6	A1X	v201835-201842	-	V
732A98	Visualisering	6	A1X	v201836-201842	2	V
732A51	Bioinformatik	6	A1X	v201845-201851	-	V
732A99	Maskininlärning	9	A1X	v201845-201851	-	O

Termin 2 (VT 2019)

Kurskod	Kursnamn	Hp	Nivå	Veckor	Block	VOF
732A55	Neuronnät och lärande system	6	A1X	v201904-201911	-	V
732A56	Webbprogrammering	6	A1X	v201904-201911	-	V
732A74	Introduktion till Python	3	A1X	v201904-201911	-	O
732A75	Avancerad data mining	6	A1X	v201904-201911	-	O
732A90	Datorintensiva statistiska metoder	6	A1X	v201904-201911	-	O
732A54	Analys av Big data	6	A1X	v201913-201921	-	O
732A91	Bayesianska metoder	6	A1X	v201913-201921	-	O
720A04	Philosophy of Science	3	A1X	v201913-201923	-	O

Termin 3 (HT 2019)

Kurskod	Kursnamn	Hp	Nivå	Veckor	Block	VOF
732A77	Artificiell intelligens	6	A1N	v201935-201944	-	V
732A62	Tidsserieanalys	6	A1X	v201936-201944	2	V
732A63	Sannolikhetsteori	6	A1X	v201936-201944	3	V
732A96	Advanced Machine Learning	6	A1X	v201936-201944	1	V
732A66	Beslutsteori	6	A1X	v201936-202003	4	V
732A76	Forskningsprojekt	6	A1N	v201936-202003	4	V
732A57	Databasteknik	6	A1X	v201945-202003	1	V
732A92	Text Mining	6	A1X	v201945-202003	2	V
732A97	Multivariata statistiska metoder	6	A1X	v201945-202003	3	V

Termin 4 (VT 2020)

Kurskod	Kursnamn	Hp	Nivå	Veckor	Block	VOF
732A64	Masteruppsats i statistik	30	A1X	v202004-202023	-	O

Hp = Högskolepoäng
VOF = Valbar / Obligatorisk / Frivillig
*Kursen läses över flera terminer