IL2230 Hårdvaruarkitekturer för djupinlärning 7,5 hp

Hardware Architectures for Deep Learning

Fastställande

Kursplanen gäller från och med HT 2021 enligt skolchefsbeslut: J-2021-0346.Beslutsdatum: 2021-04-15

Betygsskala

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

Utbildningsnivå

Avancerad nivå

Huvudområden

Elektroteknik, Datalogi och datateknik

Särskild behörighet

- Kunskaper i digitalteknik, 7,5 hp, motsvarande slutförd kurs IE1204.
- Kunskaper om mikroprocessorers uppbyggnad och instruktionsexekvering, 7,5 hp, motsvarande slutförd kurs IS1200.
- Kunskaper och färdigheter i programmering, 6 hp, motsvarande slutförd kurs DD1310/DD1311/DD1312/DD1314/DD1315/DD1316/DD1318/DD1331/DD100N/ID1018.
- Kunskaper i hårdvaruteknik, 7,5 hp, motsvarande slutförd kurs IS2202/IL2225/IL2236.
- Kunskaper i digital hårdvarudesign i HDL, 7,5 hp, motsvarande slutförd kurs IL1331/IL2203.

Aktivt deltagande i kursomgång vars slutexamination ännu inte är Ladokrapporterad jämställs med slutförd kurs.

Den som är registrerad anses vara aktivt deltagande.

Med slutexamination avses både ordinarie examination och det första omexaminationstillfället.

Undervisningsspråk

Undervisningsspråk anges i kurstillfällesinformationen i kurs- och programkatalogen.

Lärandemål

Efter godkänd kurs ska studenten kunna

- beskriva och förklara grundläggande neurala nätverk och djupinlärningsalgoritmer och deras relationer
- förklara och motivera designutrymmet för hårdvara för djupinlärningsalgoritmer
- välja och tillämpa en lämplig djupinlärningsalgoritm för att lösa verkliga problem med artificiell intelligens i inbyggda system
- analysera och utvärdera alternativ för programvaruimplementering för djupinlärningsalgoritmer
- föreslå och motivera en implementeringsarkitektur för tillämpningar med djupinlärning i inbyggda resursbegränsade system
- diskutera och kommentera nya hårdvaruimplementeringsarkitekturer för djupinlärning och nya hjärnliknande datorsystemarkitekturer som utnyttjar nya enheter och nya koncept

i syfte att

- förstå nödvändigheten, betydelsen och potentialen för att accelerera djupa inlärningsalgoritmer med låg strömförbrukning genom specialiserade hårdvaruarkitekturer
- diskutera, föreslå och utvärdera specialiserade hårdvaruarkitekturer för att implementera djupa inlärningsalgoritmer eller utnyttja djupa inlärningskoncept i resursbegränsade pålitliga system.

Kursinnehåll

Kursen består av två moduler. Modul I introducerar grundläggande kunskaper i maskininlärning och algoritmer för djupinlärning. Modul II fokuserar på specialiserade hårdvaruimplementeringsarkitekturer för djupinlärningsalgoritmer och nya hjärnliknande datorsystemarkitekturer. Förutom att presentera relevant informativ kunskap innehåller kursen laboratorie- och projektuppgifter för att skapa förståelse för de relaterade algoritmerna, att tillämpa algoritmerna för att ta itu med verkliga problem och att kontrastera och utvärdera alternativa implementeringsarkitekturer med avseende på prestanda, kostnad och tillförlitlighet.

Modul I: Algoritmer för djupinlärning

Modul I introducerar grundläggande maskininlärningsalgoritmer, grundläggande neurala nätverksalgoritmer och algoritmer för djupinlärning. Från ett antal maskininlärningsalgoritmer introducerar denna modul de algoritmer som linjär regression, polynomregression, logistisk regression, som är fundamentala och mest relevanta för neurala nätverk. För neurala nätverk tittar vi på perceptronen, multi-lager-perceptroner och i synnerhet back-propagationalgoritmen. Efter att ha gått igenom grunderna i traditionell statistisk inlärning, maskininlärning och neurala nätverk, diskuterar denna modul vidare exempel på djupinlärningsalgoritmer, specifikt Convolutional Neural Networks (CNN) och Recurrent Neural Networks (RNN).

Modul II: Arkitekturer specialiserade på djupinlärning

Modul II undersöker specialiserade hårdvarubaserade implementeringsarkitekturer för djupinlärningsalgoritmer. Från ett brett spektrum av potentiella hårdvaruarkitekturer presenteras designalternativen GPGPU:er, domänspecifika processorer, FPGA/ASIC-baserade acceleratorer, tillsammans med deras fördelar och nackdelar. I synnerhet kommer begränsningar och designalternativ för att använda djupinlärningalgoritmer i inbyggda resursbegränsade system att diskuteras. Dessutom kommer denna modul att diskutera nya arkitekturer inom djupinlärning för datorsystemdesign, såsom hjärninspirerad datorsystemarkitekturer. En fallstudie med analys, utvärdering och tillämpning av en djupinlärningsarkitektur kommer att genomföras.

Examination

- TEN1 - Skriftlig tentamen, 4,5 hp, betygsskala: A, B, C, D, E, FX, F
- LAB1 - Laborationer, 3,0 hp, betygsskala: P, 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.

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

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.