

# KUNSKAPSKONTROLL MACHINE LEARNING

Vad är skillnaden mellan Regressionsproblem och Klassificeringsproblem?

- Variablerna i regressionsproblem är beroende på kontinuerliga nummer som till exempel antal familjemedlemmar i ett hushåll. Variablerna i ett klassificeringsproblem är istället kategoriska som till exempel hudfärg och kön.

Vad är "supervised learning"?

- Ett övervakat lärande är då en funktion tränas genom märkt träningsdata som sen kan generalisera för nya datapunkter. Programmet tränas till att bedöma korrekthet och felaktighet och funktionen ändras för att ge ett korrekt resultat.

Vid klassificering tränas programmet med många exempel med det som ska sorteras. Sen måste lära sig att klassificera nya meddelanden efter att den lärt sig det gamla. Vid regressionsproblem lär sig programmet att förutsäga ett numeriskt värde med hjälp av prediktorer. Sen ges programmet flera exempel på prediktorer och etiketter.

Kan du ge några exempel på modeller som används och potentiella tillämpningsområden?

- Beslutsträd är en modell som används vid både klassificering och regression. En modell som är enkel att förstå på grund av att den så enkelt visualiserar information. Ett potentiellt användningsområde är beslutsfattande där noderna ställer fler frågor som leder till resultat eller där man vill få en prediktiv modellering.
- SVM (support vector machine) används för klassificering där support vektorer skiljer de olika klasserna åt och sen predikterar vilken klass nya observationer ska tillhöra. Ett potentiellt användningsområde är prediktion av sannolikheten av att en person har ett sjukdomstillstånd baserat på flera variabler.
- Logistic regression, då ett dataset är binärt och predikterar sannolikheten för om någonting ska inträffa eller ej. Exempel på användningsområde är om en person röstar eller inte. Resultatet ger ett utfall på antingen 1 eller 0. Över 50% blir svaret 1,

Vad är RMSE?

- Root mean square Error är en beräkningsmetod som visar hur ackurat en regressionslinje är. Genom att räkna ut minsta kvadrat från datapunkterna kan vi få en regressionslinje som är bra på att förutsäga nya datapunkter då de hamnar i ett medelvärde av de datapunkter som redan finns.

Fungerar för regressions problem då det är numeriskt. För att avlägsna negativa tal tar man roten ut och kvadraten ska vara så liten som möjligt vilket då betyder att datapunkten är nära regressionslinjen.

Vad är en "Confusion Matrix"?

- För binära klassificeringsproblem kan confusion matrix gärna användas för att lätt visualisera hur bra en prediktion fungerar och för att motverka obalanserad data och istället spegla populationens proportionerlighet.

En kvadrat delas in i fyra sektioner, där kan programmet antingen prediktera true och det är true då har programmet predikterat rätt, det kan vara false och programmet har predikterat false, även här har programmet predikterat rätt. Sen kan programmet prediktera true men det har egentligen varit false och false men att det egentligen varit true. Både dessa är felpredikterade. Man kan enkelt stapla upp och räkna ut hur mycket fel eller rätt en modell gör.

1. Utmaningar du haft under arbetet samt hur du hanterat dem.

- På grund av operation, sjukdom och vab har jag dessvärre kommit efter i mina studier. Tack vare inspelat material har jag kunnat ta igen de missade lektionerna på helger och kvällar. Eftersom Machine learning är ett intressant i sig och studietiden varit så precist planerade har det varken varit förvirrande eller svårt att komma ikapp. I ämnet machine learning har koden varit svårast för mig då jag inte arbetat med det förut, annat än i två tidigare kurser. Att förstå dimensioner, dimension-multiplikation, rymd samt själva modellerna har varit enkelt.

2. Vilket betyg du anser att du skall ha och varför.

- Jag önskar ödmjukt ett VG då mitt arbete uppfyller de krav som finns för VG enligt den studieplan som finns att tillgå. Jag har besvarat på frågorna i den teoretiska delen, modellerat MNIST-datasetet samt gjort en förhoppningsvis tillfredsställande rapport över denna.

3. Tips du hade "gett till dig själv" i början av kursen nu när du slutfört den.

- Våga fråga även frågor som av andra studenter kan vara uppenbara och självklara.