# Kunskapskontroll 1

Denna kunskapskontroll syftar till att vi ska börja tänka på vissa grundläggande och centrala koncept inom maskininlärning som är en central del inom AI. Oroa dig inte om du inte förstår alla koncept, tillsammans kommer vi diskutera frågorna på lektionen. Primära syftet för denna kunskapskontroll är att vi ska lära oss. Denna kunskapskontroll kan ge betygen IG/G.

Frågorna är kopplade till slides 1-3 från kursvecka 1.

## Frågor kopplat till slides från presentation 1.

1. Hur kan vi definiera maskininlärning?

*Det är en metod/algoritm som används för att göra prediktioner för en given datamängd. En maskininlärningsmodell byggs av en övervakad maskininlärningsalgoritm och använder beräkningsmetoder för att "lära sig" information direkt från data utan att förlita sig på en förutbestämd ekvation. Mer specifikt tar algoritmen en känd uppsättning indata och kända svar på den och tränar maskininlärningsmodellen att generera rimliga förutsägelser för svaret på nya tidigare osedda data.*

1. Vad kännetecknar regressionsproblem?

*Linjär regression är en statistisk maskininlärningsalgoritm som används för att beskriva en kontinuerlig variabel (utdata) som en linjär funktion av en eller flera variabler och hyperparametrar.*

1. Vad kännetecknar klassificieringsproblem?

*Klassificering är en övervakad maskininlärningsalgoritm där modellen försöker förutsäga rätt etikett för en given indata. I klassificeringen är modellen fullt upplärd med hjälp av träningsdata, sedan utvärderas den på testdata innan den används för att utföra förutsägelse på nya tidigare osedda data.*

1. Antag att vi har följande regressionsmodell: , vad kallas Y, X, och ?

*Y: utdata/beroende variabel/label (klassifiering)*

*X: indata/oberoende variabel/feature (klassifiering)*

*Beta: parametrar*

*Epsilon: felterm/stokastisk variabel*

1. Kan du ge ett exempel där följande regressionsmodell hade kunnat användas: ?

*Lön = startlön + löneutvecklingstakt\*anställda\_år + slumpvärde*

1. Hur kan vi utvidga/generalisera modellen i fråga 5?

*Man kan utöka med fler dimensioner*

*Dvs att utdatan beror på mer än en variabel tex anställningstid, utbildning, ort mm*

1. När man skall göra modellval och modellutvärdering så kan man använda ett tränings, validerings och test data. Förklara hur de olika delarna används.

*Träning(~65%): en delmängd av den totala som används för att träna modellen*

*Validering(~15%): en delmängd för att validera modellen iterativt under träning  
Test(~20%): en delmängd för att sluttesta, får EJ användas under träningsfasen*

*Om det är ett riktigt stort dataset kan man minska testkvoten till ~5% eller lägre.*

1. Om man vill köra K-fold cross validation så kan man börja med att dela in datan i träning och test set, förklara hur K-fold cross validation fungerar.

*Innan man delar in i träning och testdata så brukar man randomisera ordningen på datan. För K-fold brukar man dela upp träningsdatan i K=5..10 block. Ett av dem används för test under träning och blocket som används som test skiftas för varje gång vilket resulterar i K testresultat och man brukar ta medelvärdet av dem som ”score”*

## Frågor kopplat till slides från presentation 2.

1. Kan du ge 2 exempel på vad klassificeringsmodeller kan användas till?

* *Detektering av siffror i en bild, modellen predikterar vilken av siffrorna 0-9 (10 klasser) som finns i bilden*
* *Spam-detektering – prediktera spam/inte spam (2 klasser) baserat på innehåll/rubrik i ett mail.*

1. Hur tolkar du följande Confusion Matrix:   
   En bild som visar text, skärmbild, Teckensnitt, nummer

   Automatiskt genererad beskrivning

*Övre vänstra: True Negative – visar negativt när det är negativt dvs RÄTT  
Övre höger: False Positive -- visar positivt när det är negativt dvs FEL  
Nedre vänster: False Negative – visar negativt när det är positivt dvs FEL  
Nedre höger: True Positive – visar positivt när det är positivt dvs RÄTT*

*Siffrorna anger antalet av totalen 165, diagonalsumman är totalt antal rätt*

1. Hur definieras Precision? Hur tolkas Precision?

*Precision=TP/(TP+FP) och det anger andelen av de positiva prediktionerna som faktiskt är korrekta.*

1. Hur definieras Recall? Hur tolkas Recall?

*Recall =TP/(TP+FN) och det anger andelen av den positiva klassen som vi predikterar korrekt.*

1. Hur kan binära klassificerare användas för ”multiklass klassificierings problem”?   
   Ledning: OvO och OvR algoritmerna.

*Om man tex har tre klasser A,B och C kan man göra tre modeller som predikterar sannolikheten för att indatan tillhör klassen. Man måste med andra ord skapa en binär klassificerare för varje klass och välja den som får högst ”score”*

1. Hur funkar OvO och OvR algoritmerna?

*OvR: se föregående svar*

*OvO: med samma som ovan så måste man göra en modell för varje kombination av klasser – [A,B],[A,C],[B,C] = Nx(N-1)/2 modeller. Det blir många modeller men fördelen är att man endast måste träna på de fall som är inblandade i modellen tex A och B*

1. Vad innebär ”Multioutput-Multiclass Klassificiering”?

*Man kan prediktera flera variabler/lablar samtidigt, tex bounding box och objekttyp*

## Frågor kopplat till slides från presentation 3.

1. En bild som visar Teckensnitt, diagram, vit, text

   Automatiskt genererad beskrivningEn bild som visar Teckensnitt, skiss, vit, linje

   Automatiskt genererad beskrivningHur kan du skriva ut följande summor:

*1+2+3+4 = 10  
(1-4)^2 + (2-4)^2 + (3-4)^2 = 14*

1. En bild som visar Teckensnitt, diagram, vit, linje

   Automatiskt genererad beskrivningRoot Mean Squared Error (RMSE) är ett viktigt mått för att utvärdera regresionsmodeller:

Hur kan du tolka måttet?

*Man kan tolka det som prediktionens medelavstånd till de sanna värdena.*

1. Kan du kort förklara hur fixa men okända parametrarna , skattas med hjälp av MSE?

*Om man skriver om formeln på matrisform får man*

*A screenshot of a math problem

Description automatically generated*

*Första kolumnen i X är ett för att ta hand om bias-termen. För att lösa ut parametervektorn β kan man söka efter den vektor som minimerar felet i MSE-mening (vilken är en andragradsform (konvex) dvs minimum är då derivatan är noll). Om man deriverar MSE ekvationen och väljer nollpunkter kan man estimera parametervektorn som en funktion av indata och utdata enbart:  
[(p+1) x 1]=[(p+1) x n][n x (p+1)] [(p+1) x n][n x 1]*

*OBS! Detta kan mycket väl ge en lokalt minima och för att vara säker att hitta ett globalt måste man använda andraderivatan men det ligger utanför scopet på denna frågan.*

1. I frågan ovan, varför används MSE och inte RMSE? Vad är skillnaden?

*MSE ger större påverkan för stora fel, en fördel med att använda RMSE kan vara att felet har samma enhet som datan vilket gör den intiutivt lättare att förstå.*

1. Vad är intuitionen bakom ”Gradient Descent”, vad används det till?   
   *Om man löser estimeringen enligt ovan kan det bli enormt stora matrisekvationer (invertering mm) vilket kräver enorm processorkraft om man har mycket data. En variant är att söka minimum av loss-funktionen iterativt. Det gör man genom att utvärdera felet för ett subset av datan och sedan korrigera parametervektorn så att felet minskas. Stegstorlek och riktning styr hur snabbt och korrekt algoritmen konvergerar. Gradient Descent flyttar sig negativt derivatan av* *loss-funktionen. Det finns en risk för självsvängning om man tar för stora steg så att man passerar nollderivatan och kommer högre upp på den motsatta sidan av kurvan.*
2. Vad är en parameter? Vad är en hyperparameter? Vad heter dem två koncepten i Scikit-learn?

*Parameter: En parameter som lärs från träningen. ScikitLearn: parameter  
Hyperparameter: En parameter som används för att styra inlärningen. ScikitLearn: model parameter*

1. Vad är Bias Variance Trade-off?

*Det beskriver förhållandet mellan en modells komplexitet, noggrannhet och förmågan att prediktera tidigare osedda data. Det är inte säkert att modellen presterar bättre för att den är mer komplex. Bias-fel (underfitting) och variance-fel (overfitting) är två motparter och det är svårt att hitta en kompromiss mellan dem.*

1. Varför är mer komplexa modeller inte alltid bättre?

*Overfitting gör att den kan få väldigt bra resultat på en typ av data men kommer det ny osedd data får den problem. Det är oftast att föredra en generaliserad modell.*

1. Vad är intuitionen bakom att regularisera en modell? Har du tre exempel på regulariserade regressionsmodeller?

*Att regularisera en modell innebär att man ”begränsar den”, d.v.s. flexibiliteten minskar. Detta leder till högre bias men lägre varians dvs mer generell modell.  
Egentligen räcker det med ElasticNet eftersom den motsvarar både Ridge och Lasso och man styr det med alpha-parametern 0=Ridge (mer variance), 1= Lasso (mer bias), allt emellan är en kompromiss.*

1. Varför är namnet för ”Logistisk Regression” missvisande?

*Den används för klassificering, inte regression.*

1. På slide 44 används ”Logistic function” som har värden mellan 0-1, vad är syftet med detta och hur kan det tolkas? Ledning: Sannolikhets-tolkning.

*Syftet är delvis att mjuka upp övergången mellan noll och ett och begränsa utdatan [0..1]. Den har också en trevlig egenskap hos derivatan g’(x)=g(x)(1-g(x)). Om värdet är större än eller lika med 0.5 så motsvarar det en etta/positiv dvs den tillhör klassen, mindre än motsvarar 0 dvs den tillhör inte klassen.*