# Instuderingsfrågor

Frågorna kommer diskuteras i grupper och därefter diskuteras i helklass. Dessa frågor skall hjälpa oss i vår inlärning och du förväntas inte kunna alla svar.

## Frågor kopplat till slides från presentation 1.

1. **Hur kan vi definiera maskininlärning?** Maskininlärning handlar om förmågan att lära en maskin att automatisera processer för ny data baserat på befintlig data utan att explicit behöva programmera kodningen bakom. Eget insikter.
2. **Vad kännetecknar regressionsproblem?** Regressionsproblem involverar egenskaper som vi vill använda för att förutsäga ett kontinuerligt målvärde.   
   Beroende variabel kontinuerligt. Förutsäger.
3. **Vad kännetecknar klassificieringsproblem?**Klassificeringsproblem involverar förutsägelse av en kategori som en given uppsättning datafunktioner tillhör baserat på **diskreta obervationer** eller binära etiketter.  
   Fler klasser som emails med churn eller inte churn, kön,etc.
4. **Antag att vi har följande regressionsmodell: , vad kallas Y, X, och ?**Y är den beroende variabel som vi vill få baserard på X variabel dessa är oberoende features som vi använder för att predikera Y. är en parameter (en konstant) som definieras - de lär oss från datan - och bestämmer bästa linjen för att representera sambandet mellan X och Y värden. representerar kostnadsfunktionens fel eller hur mycket Y-värdena skiljer sig från X-värdena.
5. **Kan du ge ett exempel där följande regressionsmodell hade kunnat användas: ?** Denna modell kan användas för att förutsäga människors längd baserat på deras kön , där Y är det kontinuerliga längdvärdet som beror på de oberoende X-värdena för kön, man eller kvinna.
6. **Hur kan vi utvidga/generalisera modellen i fråga 5?** Vi kan hjälpa till att generalisera modellen i fråga 5 genom att lägga till fler data, parametrar och oberoende variabler, i synnerhet relevanta data.
7. **När man skall göra modellval och modellutvärdering så kan man använda ett tränings, validerings och test data. Förklara hur de olika delarna används.** När de initiala hyperparameterinställningarna har ställts in används träning för att träna olika modeller, validering används för att utvärdera den bästa modellen och hyperparameterinställningarna och testning används för att testa denna bästa modell på osedda data för att säkerställa att den generaliseras väl.
8. **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.** K-fold korsvalidering fungerar genom att dela upp träningsuppsättningen i lager eller veck, ofta 5 eller 10, och sedan efter att ha randomiserat data, itererar du genom varje veck och håller 1 veck som pseudotestdata för att utvärdera och de andra 4 (K - 1) för att träna och du tar sedan genomsnittet av varje av dessa, säg 5, prestationsresultat som modellens slutliga poäng.

## Frågor kopplat till slides från presentation 2.

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

   Automatiskt genererad beskrivning
2. Hur definieras Precision? Hur tolkas Precision?
3. Hur definieras Recall? Hur tolkas Recall?

## 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:
2. 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?

1. Vad är en parameter? Vad är en hyperparameter?
2. Vad är intuitionen bakom att regularisera en modell? Har du ett exempel på en regulariserad regressionsmodell?