

1.1. Vad är skillnaden mellan Regressionsproblem och Klassificeringsproblem?

Regression problem är när man försöker förutspå utdata baserat på indata som visualiseras genom en regressionslinje. Det kan vara att förutse nästkommande värde som till exempel att förutspå temperaturen för nästa vecka.

Klassificerings problem handlar istället om att finna en variabel/klass/kategori som en observation tillhör. Tex ett spammail som då inte tillhöra samma kategori mail av dem man vill ha i sin inkorg.

1.2. Vad är "supervised learning"?

Supervised learning är när man har både indata av ex en katt, och även utdata i formen av namnet en katt, eller rentav olika raser av den katten. Dvs, man har både indata och utdata känt innan, som man lär modellen med. På detta sätt ska modellen kunna lösa vad indata i form av andra katter kunna antingen definieras som en katt eller om modellen är mer avancerad vilken ras av katten, dvs tidigare osedd indata ska kunna generera korrekt utdata.

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

Linjär regression och kan användas inom ekonomi för att tex förutse huspriser.

Logistisk regression som används vid binär klassificering, tex för titta på en symptom och se om en person har en sjukdom eller inte.

Decision trees and Random Forest som kan användas tex inom spel för att skapa "non playable characters" (npc) artificiella beteende.

Neurala nätverk och Deep learning, som används vid komplexa uppgifter som tex bildigenkänning.

Vad är RMSE?

RMSE(Root Mean Square Error) är en statistisk måttstock som används för att mäta skillnaden mellan det förutspådda värdet och det faktiska värdet i ett regressions problem. Ett lägre RMSE värde visar bättre passform till datan.

Vad är en "Confusion Matrix"?

Confusion Matrix en tabell som används för att visualisera antalet hur väl en klassificerings algoritm fungerat. Det som jämförs är det sanna värdena mot de predikterade värdena.

I ett binärt system:

	Predikterat: Nej	Predikterat: Ja
Sant Värde: Nej	TN	FP
Sant Värde: Ja	FN	TP

TP står för True positive och visar det predikterat värde Ja har det sanna värdet Ja.

TN står för True negative och visar det predikterade värdet Nej och det sanna värdet är Nej.

FP står för False positive och visar det predikterad värdet Ja, men sanna värdet är Nej.

FN står för False negative och visar det predikterat värdet Nej och det sanna värdet är Ja.

I ett multi klassificerings problem

	Predikterat A	Predikterat B	Predikterat C
Sant Värde A	10	2	1
Sant Värde B	3	15	0
Sant Värde C	4	0	18

Predikterat värde A och Sant Värde A är i detta exempel 10. Så 10 värden som är predikterat som A är också sanna för A.

I andra kolumnen av Predicted B, första raden så visar den 2. Så detta är det predikterade värdet B men det sanna värdet är A.