

Det egna projektet, steg 1

Du skall nu skapa ett eget projekt, där du använder maskininlärning/AI till att lösa ett på förhand givet problem.

Du väljer själv vilket problem du vill försöka lösa. Här finns inga egentliga begränsningar.

Exempel på uppgifter är:

- Identifiera saker/djur/människor i bilder
- Tolka handskriven text
- Förutsäga väder
- Förutsäga börskurser
- Tolka patientdata för att förutsäga risk för vissa sjukdomar
- Identifiera "fake news"

Börja med att i text beskriva problemet och vad det är du vill åstadkomma. Nästa steg är att försöka hitta data som kan hjälpa dig lösa ditt problem. Om du inte kommer över lämpligt data får du tänka om. Kan du göra en variant på ditt problem som kan lösas med det data du hittat? Om inte, definiera ett nytt problem och försök igen.

Tänk på att datat kan komma från olika källor. Om du hämtar data från olika källor måste du undersöka hur kompatibelt det är.

Lär känna ditt data.

- Är det komplett?
- Har du null-värden?
- Har du extrema värden?
- Vilka datatyper har datat?
- Vilka fält i ditt data vill du använda dig av?
- Hur kan du konvertera alla fält du vill använda till ett numeriskt format?

Typ av problem

Vilken typ av problem är det du försöker lösa? Är det ett klassificerings- eller regressionsproblem? Är ditt data labeled (alltså har du färdiga svar att mata din modell med vid träning). Om inte, hur skall du få det? Skall du själv skapa labels eller vill du försöka att ta hjälp av unsupervised learning (fungerar bra på klassificeringsproblem men inte regressionsproblem).

Vill du i stället arbeta med Reinforcement Learning och deep neural networks? Du får du fundera vad du vill uppnå och vilket data du vill mata ditt nätverk med. Hur skall det vara strukturerat? Hur skall ditt belönings/bestraffningssystem se ut. Hur skall du tolka det output du får från nätverket? Vilken träningsplattform vill du använda för att träna ditt nätverk?

Dokumentera

Detta är det du skall göra i steg 1. Du skall alltså inte börja implementera någon lösning ännu, utan undersöka förutsättningarna för ditt projekt. Skapa ett dokument där du beskriver problemet och den lösning (enligt frågeställningarna ovan) som du kommit fram till. Om du använder dataset, beskriv det data du hittat. Hur många rader har du? Hur är kvaliteten på datat.

Du kommer lämna in detta dokument i nästa inlämning. Du kan också, om du vill, lämna in Jupyter Notebooks som visar hur du har undersökt ditt data (om du inte skall arbeta med Reinforcement Learning).