

Laboration 3

I denna laboration skall vi undersöka en *linjär klassificering*. Denna laboration skall lämnas in som en `.py` fil (din kod), en `README.md` fil som beskriver vad ditt program gör och hur man använder det samt en `labelled_data.csv` fil (svaret på uppgiften). För VG skall även en `report.ipynb` fil finnas i repot.

G

a) Börja med att läsa in filen `unlabelled_data.csv` och plotta innehållet i en `scatterplot`. Du behöver inte lämna in denna del – detta är för att du skall kunna göra nästa steg. Använd antingen en jupyter notebook eller ett python skript, vilket du tycker är enklast.

b) Dra en rak linje genom figuren som delar punkterna så jämt som möjligt. Bestäm denna linjes ekvation på formen $y = kx + m$.

Ledning: Tänk på vad värdena måste vara när x är 0 eller y är 0.

c) Implementera en funktion som räknar ut om en punkt ligger till höger om/o-vanför eller vänster om/nedänför linjen. Först nu skriver du kod som du skall lämna in. Skriv denna funktion i en `.py`-fil.

Ledning: Om y värdet för en punkt är större än linjens värde för motsvarande x , vad betyder det?

d) Skapa ett program som använder funktionen från förra uppgiften för att klassifiera alla punkterna i `unlabelled_data.csv` filen.

e) Skriv till en fil `labelled_data.csv` där du lagt till en kolumn med 0 om punkten är till vänster om linjen och 1 annars.

f) Visa ett fönster med en graf där punkterna, deras klass och din linje alla visas.

VG

För VG skall du även lämna in en väldigt kort rapport som en jupyter notebook. Rapporten skall innehålla följande:

i) Visa datan i en graf tillsammans med följande linjer utöver den du själv hittat:

$$\begin{aligned}f(x) &= -0.489x \\g(x) &= -2x + 0.16 \\h(x) &= -1.08x - 0.1\end{aligned}$$

ii) Använd även dessa nya funktioner för att klassifiera datan. Är det några skillnader?

iii) Resonera kring hurvida det finns någon anledning att föredra någon viss linje eller inte. Hur många sådana här linjer kan vi hitta?

Skriv i löpande text som markdown-celler. Se till att koden i python-celler är kortfattad och hör ihop med texten runtomkring. Rapporten skall vara resonerande men behöver inte några källhanvisningar. Visa din tankegång med text, kod och grafer. Antag att läsaren kan allt du kan. Återanvänd kod där möjligt.