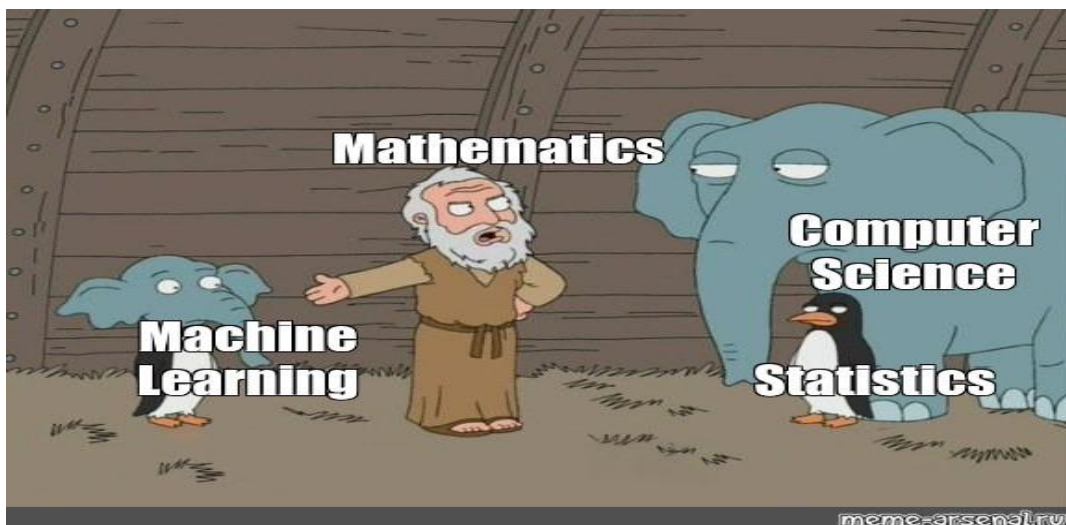


Machine Learning



=> امتی نستخدم او نلجاء ل ML ؟

- ☐ لو روحنا لل data science وهي معرفتش تديني قرار أو تخليني أفهم الداتا
- ☐ احنا في مصر اصلا بنقف عند ال analysis لأن احنا مينعرفش ن formulate ال problem لكود

=> ال ML بيحل أنواع من ال problems :

1. Classification
2. Regression
3. Clustering
4. Generation

=> ايه هو ال Classification ؟

- ☐ معايا صورة وعاوزه أعرف هي قطه ولا كلب ولا حصان مثلا , يعني عندي شوية اختيارات محددة وثابته وهو بيقولي ايه الصح فيهم , يعني احنا عارفين كل الأوبشنز الممكنه بس مش عارفين مين الصح فيهم
- ☐ في ال classification كل اما تزيد عدد ال classes أو الأوبشنز كل ام ال accuracy بتاعتي بتقل أو خلينا نقول دقة ال model في اختيار الاختيار الصح هتقل
- ☐ فيه منه نوعين :

1. Binary classification
2. Categorical classification

=< ال ML بيطلع ال classification إزاي ؟

□ بيطلعها بال probabilities يعني بيقلولي الأختيار ده صح بنسبة كام في المية

=< طيب انا هستفيد بيه في ايه ؟

□ لو انا معايا مودل بيعمل binary classification بمعنى بيفرق ما بين حاجتين yes or NO , قطة وكلب , ولد وبنت وهكذا

□ لو انا أدبت للمودل صورة قطة وهو اصلا بيفرق ما بين القطة والكلب فهو قالي إنها قطة فعلا بس بنسبة 51% فده معناه انه جاوب صح بس بالميت اوي يعني لو جربنا نديله صورة قطة فيها تشابه بالكلب مش هيعرفها

□ فالكلام ده انا مش علوزه

=< طب أمتي المودل يكون مش واثق ودي تكون حاجه كويسه ؟

□ لو انا ك Expert في ال domain الي المودل شغال عليه ده مش واثق اصلا وده لإن الحاجات الي بيفرق المودل ما بينهم متشابهين بدرجة كبيرة

=< يبقى احنت كده عندنا مشكلتين في ال classification :

1. كل اما يزيد عدد ال classes ال accuracy احتمال كبير انها تقل

2. إن المودل بيطلع ال output بتاعه probabilities

=< ال Regression :

□ عكس ال classification وهو إني بب predict ال value بس ف continuous range زي درجة الحرارة لما بتيجي تتغير مش بتتغير بأرقام ثابتة بل بتزيد او بتقل بأرقام fractions

=< ال supervised learning :

□ Supervised معناها إني اقدر اتحكم في عملية ال learning , يعني انا اصلا معايا الإجابات الصح وبخلي المودل يتعلم ويطلعلي إجابته واقرنها بالصح

=< ال clustering من ضمن ال unsupervised learning :

□ ال unsupervised فكرة إني معييش الإجابات

- ال clustering هو فكرة إني معايا مجموعه من الداتا انا بقسمها ل groups
- لو انا معايا فصل وعاوزه اقسم الكلاب الي فيه لمجموعات ومعايا كذا تقسيمه بس انا معرفش مين أحسن تقسيمه يبقى انا عارف ومتأكد إن في تقسيمه صح بس انا معرفش ايه هي
- لما باجي ا cluster عندي 2 challenges :
 1. على اي اساس هقسم الداتا الي معايا
 2. وهقسمهم كام group
- هقسم على اي اساس دي بتتحدد عن طريق هو انا اصلا متاح لي التقسيمه عامله ازاك business
- وانا اصلا عاوز أقسم ليه ممكن ا cluster كذا مره بتقسيمات مختلفه وعلى اساس مختلفه لكذا غرض
- لو انا معنديش domain knowledge فمش هعرف ايه احسن تقسيمه ومش هعرف ا cluster
- مشكلة هنقسم كام جروب دي limited بال resources وبال testing وبال time , ممكن أقسم 3 جروب او 4 على حسب

=> دي انواع ال problems المختلفه الي ال AI بيحلها كل problem من دول ليها algorithm كل algorithm بيتحدد امتى استخدمه على حسب نوع الداتا الي معايا

=> في اكثر من library بنعمل بيها ML :

1. Sklearn

2. Pyspark

3. Ml clear

4. H2o

5. T-pot

=> اغلب الناس بتستخدم Sklearn وده لأن ال syntax بتاعها متشابه جدا بس ليها limitation وهي انها مينفعش تشتغل على data set كبيره .

=> لو معايا data set كبيره هستخدم pyspark.

ml clear ميزتها انها بعد ما بتعمل المودل بتعمله **automatic** **deployment** و **visualisation** للمودل وال **accuracy** بتاعته مع بعض .

h2o ,t-pot انت بتديهم الداتا وهما بيحربوا كل ال **models** الممكنه وي **save** المودل الي عنده أعلى **accuracy** ويعمله **automatic** **deployment** و **visualisation** .

=> في حاجه اسمها ال self supervised learning :
☐ وهو أن ال model بيعلم نفسه بنفسه "أغلب استخدامه في ال games"

=> في ال semi-supervised learning :
☐ وهو فكرة إن معاك شوية داتا **labeled** وشوية مش **labeled** والداتا الي مش **labeled** اكتر بكثير يعني ايه مش **labeled** يعني منعرفلهاش ملامح فاحنا بنستخدم الداتا ال **labeled** عشان نعرف بقيت الداتا ☐ وهو ده فعليا الي بيحصل في الحقيقه لأن اغلب الداتا الي موجوده ملهاش class

=> ال Generation :
☐ هو فكرة اني عاوزه ا generate الداتا يعني انا مش عاوز output لا انا عاوز داتا

☐ طب ليه مضطرين إن احنا ن generate data ؟
☐ ممكن نعمل data augmentation , ممكن يكون عندي الداتا قليله , ممكن نعمل داتا جديده based على داتا قديمه معايا وده ال بيعمله ال GAN algorithm .
☐ في مجال advanced اكثر من ال generation اسمه neural evaluation

<= ال neural evaluation :

□ هو نفس فكرة التطور الجيني وهو يستخدم في ال genetic algorithm مع ال AI تحديدا مع ال DP فاحنا لما بناخذ مودل بنعتبر إن معايا 500 او 600 مودل بيحلوا نفس ال problem بناخذ الأفضل منهم ون retrain ال models دي تاني ونمكس ما بين ال parameters بتاعتهم يعني بناخذ ال parameters الحلوه من هنا على الحلوه من هناك وناخذ generation أقل شويه في العدد ولكن ال models دي بتجيب اعلى accuracy وتكرر الصورة دي مع كل generation لجد ما تطلع ب super powerful models وال accuracy بتاعتها عاليه جدا .

□ الي قدر يطبق الكلام ده لحد دلوقتي هي شركة NVIDIA .

<= Types of models :

Statistical model □

Non-statistical model □

<= Statistical model : مبنيه على شوية قواعد و rules ثابتة وواضحه

<= Non-statistical model : ال models الي مش شغاله بب pattern وطريقة ثابتة في إنها تختار ال features الصح عشان ت predict بيها .

يا ابن آدم إنما أنت أيام .. إذا ذهب يومك ذهب بعضك

الحسن البصري