

Post 1

ننتقل من ال Regression على ال Classification وخلينا نفكر فى Image classification للأرقام من 0-9 لو رجعنا لموضوع ال Cross Validation واخترت مثلا 3-folds وكل منهم أنا عارف انه بيعمل train لجزء من ال fold و validation للجزء الباقي وهكذا لكل fold لحد ما الموديل يكون عمل Train على كل ال Data وفى نفس الوقت عمل Validation على كل ال Data المهم هنا والى كنا اكلمنا عليه قبل كده هو التأكد فى التقسيمات ديه إنى كل صور الأرقام تكون موجوده فى كل Fold مش بس كده ال Fold الواحد إنتا محتاج إنك تتأكد هل ال Train and Validation set فيهم كل الصور لكل الأرقام بمعنى ميجيش يكون فى ال Train فيه صور الأرقام من 0-7 ونيجى نلاقى فى ال Validation موجود كل الصور من 0-9 فهنا فى ال Fold ده جزء ال Train مشفش أصلا 8 و 9 وهكذا ، وكمان أخلى بالى من النسب لأن بعد كده ممكن يكون فيه نسبة Error مثلا بسبب إنى مثلا الصور بتاعت رقم 5 فى الداتا سبت قليلة ، وهنا نخلى بالننا من ال shuffle الى بيحصل حاجة زى ال Stratified shuffle بيساعد فى التقسيم بنسب متقاربة لكل ال classes المختلفة، وبرضه ده بيعتمد على نوع ال data يعنى مينفعش داتا معتمدة على الترتيب الزمنى نعمل فيها shuffle لأن الترتيب هنا يفرق مثلا لو بقيس ال progress بتاع الطلاب فى الجامعة. الداتا معايا والدنيا متظبطة وقولت اعمل run للموديل وأشوف ال Accuracy فهل ده صح بالنسبة ل classification ولا المفروض اعرف الموديل كويس إزاي ده بإذن الله فى البوست الجاى ^^

https://www.linkedin.com/posts/abdelrahman-rezk_data-ml-classify-activity-6778377657328431104-G3fy

Post 2

لو جينا بصينا للنتيجتين من الموديل هانلاقيهم حققوا فوق 90% رغم انى هنا عدد ال images-5 بالنسبة لكل الصور قليل بنسبة 10% من ال images-5 ل 90% من Non-5-images طب هنا ليه حقق فوق ال 90% فى الاتنين لان احنا مجرد بنحسب Accuracy ايه الى اتصنف صح على الى متصنفش صح لكن مبخدش فى اعتبارى الحاجات الى اتصنف على اساس انها 5 او العكس ال False Positive or False Negative errors ومن هنا بدل ما احنا علطول نروح نجيب ال Accuracy او حتى عن طريق انى ال Sklearn بترجعلى ال Score علطول عن طريق مثلا Cross_val_score فى class تانى وهو بيرجع ليا ال Prediction بتاع كل instance ومنه ابتدى اشوف الطريقة الى انا عايز احسب بيها فمثلا من الحاجات الى ممكن اشوف بيها نتيجة الموديل بتاعى عشان اشوف ال error اكتر فى انهى حاجه هى ال Confusion Matrix. ال Confusion Matrix عبارة عن ببساطة 2-d array طولها زى عرضة بعدد ال classes الى فى ال output بتاعك ال Row بيعبر عن Class نفسه بينما ال Column بيعبر عن ال Predictive فلو جينا لل row الأول الى هو ال Negative class الى هو بالنسبة ليا ال Non-5-images نلاقى اول رقم ده بيعبر عن ال TN وهى الحاجات الى فعلا اتصنفت صح انها صور مش بتحتوى على رقم 5 بينما الرقم الى جمبه ده هايحبر عن الحاجات الى حصلها Confusion مع ال Non-5-images الا وهى ال images-5 معنى كده انى ال False Positive هو الحاجات الى حصلها تصنيف غلط انها images-5 ننزل لل row الى بعده وهو ال Positive class بتاعنا الى احنا عايزنه اصلا وهو ال images-5 نلاقى انى الرقم الأول فى ال row بيعبر عن الحاجات الى حصلها تصنيف غلط انها Non-5-images الا وهو ال False Negative معنى كده ايه ديه الحاجات الى حصلها Confusion مع الكلاس 5 واتصنفت غلط انها Non-5-images اخر حاجه وهو ال True Positive الا وهو الحاجات الى اتصنفت صح انها images-5 ومن هنا يطلع عندى انى اهم حاجة هى انى يكون الرقمين الى فى ال Diagonal الى هما True Positive and True Negative يكونوا أكبر ما يمكن. طبعا الكلام ده كله جميل ولكنه ممكن يلغبط + انى لو عايز اقارن موديل ببعضها محتاج رقم واحد يعبر عن هل النتيجة بالنسبة ليا كويسه ولا لا وغير كده لو عايز اقلل ال False Positive او False Negative المفروض اعمل ايه وهنا بيتدى يظهر عندنا من ال Matrix نفسها حاجتين مهمين جدا الا وهما ال Precision and Recall وده بإذن الله فى البوست الجاى ٨٨

[/https://www.linkedin.com/pulse/confusion-matrix-abdelrahman-rezk](https://www.linkedin.com/pulse/confusion-matrix-abdelrahman-rezk)

Post 3

العلاقة بين ال Precision and Recall هى علاقة عكسية والى بيؤثر فى الإيتين هو ال Threshold بتاع ال Decision function لما تزودة ببقا ال Recall هايقل وبالطبع ال Precision هايزيد ! بس مش دايمًا لكن فى النهاية هايزيد يعنى لو بصينا للصورة الى معانا هانلاقى إنى تحريك ال Threshold مثلا one step لليمين هانلاقىه خلى ال 5 الى كانت True Positive بقت False Negative وبقا بدل ما كان ال Precision 80% بقا 75% وفى نفس الوقت ال Recall ببقل دوما ولكن فى النهاية لو عليت ال Threshold تانى هالاقى ال Precision بقا 100% طب أصلا هو ال Precision ده بيعبر عن إيه ببساطة عن إتنا بنسبة كام % هانتوقع ال Positivist class بتاعك صح بينما ال Recall إتنا فعلا عملت Detect لكam Positive class من الى انتا بتدور عليهم وهنا 100% Precision بيان إنه يكون أحسن لكن ده مش صح ده بيرجع فى الآخر لل Application الى إتنا شغال عليه ولكن غالبا إتنا بتحتاج تعمل Trade-off توازن بين ال Precision and Recall وده عن طريق إنك ممكن تعملهم Plot طب أنا هاوصلهم أزاى ال sklearn بتوفر class يجبلك كل ال Precision, Recall and Threshold لكل ال Instances الى بتعملها Predict ومنه تقدر تشوف تأثيرهم بالنسبة لل Threshold ومن ال Plot تختار أنسب قيم ليهم طب هل ده برضه كافى بالنسبة ليا لو محتاج أختار Model ما ، فى الجزء الى فات ده هايساعدك تحقق Trade-off بينهم عشان لما تيجى تشوف حاجة زى ال F1-score الى هاتساعدك تقول والله ال model كويس ولا لا ده معتمد فى الأساس على إنى الاتنين يكونوا متناسبين ليه لأن ال F1-score بتحسب حاجة إسمها ال Harmonic Mean يعنى إيه ببساطة بتدى قيمة أعلى للأقل Weight معنى كده سواء ال Precision or Recall أى واحد فيهم قيمته قليلة عن التانى هاتلاقى ال F1-score قيمتها صغيرة ، ده ميمنعش إنى فى بعض ال Applications ممكن أحتاج حاجة أعلى من الثانية زى بظبط لو بعمل برنامج للأطفال مثلا بيعمل Recommendation videos ليهم ف أنا عايز الفيديو الى يروح للطفل يكون مناسب جدا وميكنش فيه اى حاجة خارجة أو سيئة هنا هاكون محتاج يكون ال Precision عالى زى لما عليت ال Threshold فقدت 5 اه زى ما فى الصورة بس أنا ضامن كل ما عليت ال Threshold ال Precision هايعلى وال Safe videos هاتكون أضمن إنها هى الى تطلع رغم انى بعضهم ممكن افتقده زى ال 5 كده بظبط بس ده هايبكون أحسن بالنسبة للبرنامج بتاعى .. طب هل فى طريقة ثانية لتحقيق ال Trade-off اه ونشوفها إن شاء الله المرة الجاية ^{٨٨}

<https://www.linkedin.com/pulse/precision-recall-trade-off-abdelr/ahman-rezk>

Post 4

اتكلمنا عن ال Precision وشوفنا إيه التأثيرات لما ال Threshold يزيد ولقينا في بعض الأحيان ال Precision بيقل لكن في النهاية هانلاقية إنه زاد في نفس الوقت ال Recall بيقل طيب ال Recall زى ما كنا بنقول هو عبارة عن أنا عملت كام Detect للحاجة الى انا عايزها طب ده ينفع في ايه حاجة يكون فيها مشكلة لو عملت detect على الفاضى بمعنى لو مثلا بعمل حاجة بتحاول إنها تشوف هل شخص ما سرق حاجة ولا لا هنا أنا محتاج الى اعمله Detect يكون فعلا سرق حاجة على عكس لو شغال على App ليه علاقة بالطب أنا مش مهم عندي بعض الناس الى معندهم مش حاجه اعتبرهم مرضى فهنا هاحتاج اعلى ال Precision على حساب ال Recall طيب ده ليه اسم تانى وهو ال Sensitivity ومتعلق أكثر بال True Positive Rate الى هو ال Positive class بتاعى أنا جيت منه اد ايه بالنسبة لل False Negative الى هو الحاجات حصلها Confusion معاه وروحت صنفقتها إنها مش تبع ال Positive class بتاعى عشان ققدر اشوف ال Recall بصورة أوضح في عندي ال ROC class بيرجعلى ال True Positive Rate and False Positive Rate وال Thresholds لكل ال instances ومنه ققدر اعمل plot واشوف عند كام Recall ابتدا يظهر عندي False positive classification ومنه ققدر اختار افضل قيمة لل Recall والى بتكون قبل ما يخرج من محور الصادات زى ما في ال Graph دية كلها حاجات بتساعدنى في ال Binary classification لكنها بتفقد تأثيرها مع ال Multi classification problems وده إن شاء الله البوست الجاى^{٨٨}.

https://www.linkedin.com/posts/abdelrahman-rezk_ml-classification-precision-activity-6779471971211100161-JvvP

Post 5

نرجع لل Confusion Matrix ونطبق على Binary problem

خلينا الأول نبص للصورة ونعتبر نفسنا في Image classification للارقام بس مش كل الأرقام لا صور الرقم 5 ده كلاس وكل الصور الثانية ده كلاس تاني
ال row بيمثل في matrix الى Actual Class while column is predicative class

ال Negative class في الصورة هو ال Row الأول وال Positive class هو ال Row الثاني

بمعنى ايه ال Positive ده الى انا بحاول اعمله Predict الى هو في ال Labels بتاعتى مخليه ب 1 ب True وال Negative الى مديه 0 الى هو False
حلو اول رقم وهو 53057 ده اسمه ال True Negative وده بيعمل ايه ده ببساطة هو ال Correctly classified as Non-5 بس كده الثاني المقابل ليه في ال Diagonal هو ال True Positive وهو ال Correctly classified as 5 والاتنين دول او ال Diagonal في الماتركس ديه اهم حاجه يكون فيه اعلى الارقام
نيجى بقا للرقمين التانيين الى شغلى كله بيكون عليهم عشان ققلهم قدر الامكان وهما الاول 1522 وده ال False Positive وده ببساطة لو طلعتا فوق قلت ال row الاول بتاع ال Negative class معنى كده ايه ديه الحاجات الى حصلها Confused مع ال 5 ليه لان ال Negative class هو ال Non-5 طب دول مالهم wrongly classified as 5

وال 1325 هو ال False Negative وده الى هو ال row الثاني الى هو ال Positive class وده ماله الى حصله Confused مع ال Non-5 ليه لان ال positive class هو بالنسبة ليا ال 5 طب وده ايه دول ال wrongly classified as non-5

https://www.linkedin.com/posts/abdelrahman-rezk_ml-classification-handson-activity-6783680219434815488-IJ4H