

# 1- Post 1

مشكلة زي بتاعت ال Overfitting or underfitting ممكن تكون متعرفه بالنسبه ليك قبل ما تشتغل اصلا فى المرحلة بتاعت ال Pre processing الى بتعملها على الداتا لما تلاقى انى فى مثلا outliers كثير او تلاقى attributes or features مقارنة لعدد ال instances(dataset rows) فهنا لازم تعالج حاجة زي كده قبلها او تجرب pipelines مختلفة ممكن جدا لو بتشتغل على Linear Regression problem يكون سبب من ضمن الاسباب هو انك مش عامل Features Scaling للداتا بتاعتك وهو انك تخلى ال Input range للارقام كله مثلا من 0 ل 1 ممكن تقلل في Hyperparameters وممكن تزود عدد ال Training Data كلها محاولات ممكن تتعمل

نفس الموضوع بالنسبة لموضوع ال Underfitting بيبكون العكس حتى فى حل المشكلة لو مثلا هناك كنت محتاج تستخدم موديل مثلا Linear بدل مثلا Non linear one العكس مع ال Underfitting انتا محتاج موديل اعقد عشان يقدر يعمل Detect لل features الى ممكن يكون بينها تقارب زي مثلا موضوع صور الارقام رقم زي ال 8 يشبه كثير جدا رقم 6 فى انى هنا مثلا فيه دائرتين بينما التانى هى 1 فقط فلما الموديل يتعمل محتاج يعمل Detect لل matrix كويس جدا لانها فى الاخر هاتختلف فى جزء بسيط جدا وهكذا العلاقة بينهم عكسية الى حد ما

# 2- Post 2

الفكرة مش فى انى الموديل يعمل run كويس على ال training data وده بيقدونا للفرق بين مصطلحين مهمين وهو ال instance base and model base learning وهو انى الاول بيتعلم عن طريق التشابه يعنى مثلا لو مثلا الكلمات بين ميل جديد ميلات قديمة كبير يحصل انه هايقول عليها spam وده صح بس هنا لو نفس الكلمات جت بس الشخص الى بيتكتب الميل غير عرض الكلمات زي ايه مثلا هو كان متعود انى الكلمة بتيحي كده

u 4

فلو حصل واتغيرت الكلمة مثلا ل for u الى هو المفروض انى برضه ده spam فمش هايعرف يعمل classification لانه معتمد على التشابه على عكس بقا ال model base هو بيتعلم عن طريق ال weights الى بتتغير مع تغير ال inputs وطبعا المتحكم فى التغير ده يكون قوى ولا بسيط او يحصل توازن هو ال learning rate

# 3- Post 3

انا كنت فاكرو موضوع ال cross validation ده بستخدمه لاني اعمل test للموديل بتاع بس طلع فيه فرق كبير بين كلمتين Validtion and Test الفكرة فى موضوع ال Cross-Validation هو انى يستخدم ال Train set بتاعتي ويروح يقسمها لجزء يعمل عليه Train والتانى يعمل عليه Validate فهنا بدل ما كنت بروح ققسم الداتا لجزئين جزء Train وجزء ال Test لا بقا عندي Train, Validation and Test ومفهوم ال Cross نفسه هو انى فى نفس الوقت بقدر اعمل Train على كل ال Training set ويعمل Validation على كل ال Train set برضه طب ازاى لما نيجى نقسم ال folds بتاعتنا ل مثلا 3 هو بيروح يجيبك جزء لل Train وجزء تانى لل Validation وبعدها فى ال fold الى بعديه يروح ياخذ من جزء Train ده جزء

جديد يعمل عليه Validation وفي اخر fold هايتبقا اخر جزء مخدهوش من ال Train فياخذ يعمل عليه Validation وخلص ويرجلك ب 3 قيم لكل Validation في كل fold بعد ما خلاص حسيت اني الموديل ده ينفع انك تستخدمه وراضى بالنتيجة ساعتها تشوف بقا الموديل بتاعك على Test set هايعمل ايه

## 4- Post 4

الفكرة مش في اني الاقي داتا او يكون في داتا كثير وخلص اهم حاجة زى الموديل بظبط الى هاترنه على الداتا هي الداتا نفسها وهنا كان في الكتاب بيناقش اكثر من حاجة الاولى انك معندكش داتا كفاية وهنا الكاتب ضرب مثال جميل وهو اني الموديل مش زى الطفل الطفل ممكن والدته لما ياخده معاه مشوار يساله ايه ده يقوله تفاح مره والتانية والتالته الطفل خلاص بيتعلم اني ده تفاح بلس ممكن لون التفاح نفس الطفل يحتاج يشوف كل لون ويعرف اني كل لون من دول والشكل ده برضه تفاح ومن هنا الموديل عامل زى الطفل في انه محتاج يشوف كل انواع والوان التفاح في الداتا بتاعتك + انه محتاج يشوف عدد كبير جدا من التفاح مش مجرد كام مثال زى حالة الطفل والجزئية الثانية بتاعت انه يشوف كل الانواع ديه بتقودنا للنوع التاني وهو ال Non Representative Data وهو اني محتاج يكون الداتا بتاعتك بتمثل كل الانواع من التفاح ويكون التمثيل ده متقارب في النسبه بمعنى ميكونش في تحيز ناحية نوع من الانواع على حساب الانواع الاخرى مش كده وبس كل تفاحه بببقا ليها شوية ميزات فممكن يكون في كثير من التفاح الميزات الى بتمثله مش موجوده او عباره عن nan values وده الى بنسمية Poor-Quality Data كمان في الجزئية ديه محتاج اخلى بالي من حاجة مهمه وهي ال outliers هو امثلة ممكن تخلى الموديل يعمل overfitting وهما كام مجرد مثلا 100 مثال في ملايين الصفوف وهنا على حسب المشكلة ممكن تدرس ال 100 مثال دول على جمب لوحدهم